Effects of Methyl-Group Metabolism and Lifestyle Factors on Genome-Wide DNA Methylation

Pooja Rajendra Mandaviya

ACKNOWLEDGEMENTS
The work presented in this thesis was conducted at the Department of Internal Medicine and Department of Clinical Chemistry of the Erasmus University Medical Center, Rotterdam, the Netherlands.

This Study is funded by the Departments of Internal Medicine and Clinical Chemistry of the Erasmus Medical Center and Erasmus University, Rotterdam, the Netherlands Organization for Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam.

This work was done within the framework of the Biobank-Based Integrative Omics Studies (BIOS) Consortium funded by BBMRI-NL, a research infrastructure financed by the Dutch government (NWO 184.021.007).

The infrastructure for the CHARGE consortium is supported in part by the National Heart, Lung, and Blood Institute grant R01HL105756.

The publication of this thesis was financially supported in part by the Erasmus University Rotterdam, the Netherlands.

Layout:	Pooja R. Mandaviya
Cover design:	Cheyenne Veen, Optima Grafische Communicatie and Pooja R. Mandaviya
Printing:	Optima Grafische Communicatie

ISBN: 978-94-6361-112-1
© Pooja Rajendra Mandaviya, 2018

No part of this thesis may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, without prior written permission from the author or, when appropriate, from the publisher of the manuscript.

Effects of Methyl-Group Metabolism and Lifestyle Factors on Genome-Wide DNA Methylation

Effecten van Methyl-Groep Metabolisme en Leefstijlfactoren op Genoom-Brede DNA-Methylatie

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof.dr. H.A.P. Pols

and in accordance with the decision of the Doctorate Board.

The public defence shall be held on
Tuesday 5 June 2018 at 15:30 hrs

by

Pooja Rajendra Mandaviya
born in Porbandar, Gujarat, India

Erasmus University Rotterdam

PROMOTION COMMITTEE

Promoters:	Prof. Dr. A.G. Uitterlinden Prof. Dr. J. Lindemans
Inner Committee:	Prof. Dr. C.L. Relton Prof. Dr. J.H. Gribnau Prof. Dr. M.A. Ikram
Co-promoters:	Dr. J.B.J. van Meurs Dr. S.G. Heil

Paranymphs:	C.K.A. Fleming H.R. Mandaviya

To my parents Rajendra & Sunita

To my grandmother Sushila

To my late uncle Hasmukh

CONTENT

List of abbreviations 9

Part A **Introduction**

Chapter 1 General introduction and outline of thesis 13

Chapter 2 Homocysteine and DNA methylation: a review of animal and human literature (Molecular genetics and metabolism, 2014) 31

Chapter 3 Nutrients and DNA methylation across the life course: a systematic review (To be submitted) 67

Part B **Homocysteine and DNA methylation**

Chapter 4 Homocysteine levels associate with subtle changes in leukocyte DNA methylation: an epigenome-wide analysis (Epigenomics, 2017) 133

Chapter 5 Genetically defined elevated homocysteine levels do not result in widespread changes of DNA methylation in leukocytes (PLoS One, 2017) 163

Chapter 6 Interaction between plasma homocysteine and the *MTHFR c.677C>T* polymorphism is associated with site-specific changes in DNA methylation in humans (Submitted for publication) 185

Part C **Nutrition, lifestyle and DNA methylation**

Chapter 7 Association of dietary folate and vitamin B12 intake with genome-wide DNA methylation; a large scale epigenome-wide association analysis in 5,841 individuals (Submitted for publication) 209

Chapter 8 Epigenetic Signatures of Cigarette Smoking (Circulation. Cardiovascular genetics, 2016) 237

Part D **Discussion and summary**

Chapter 9 General discussion 265

Chapter 10 Summary 285

Part E **Appendices**

List of publications 293

PhD portfolio 297

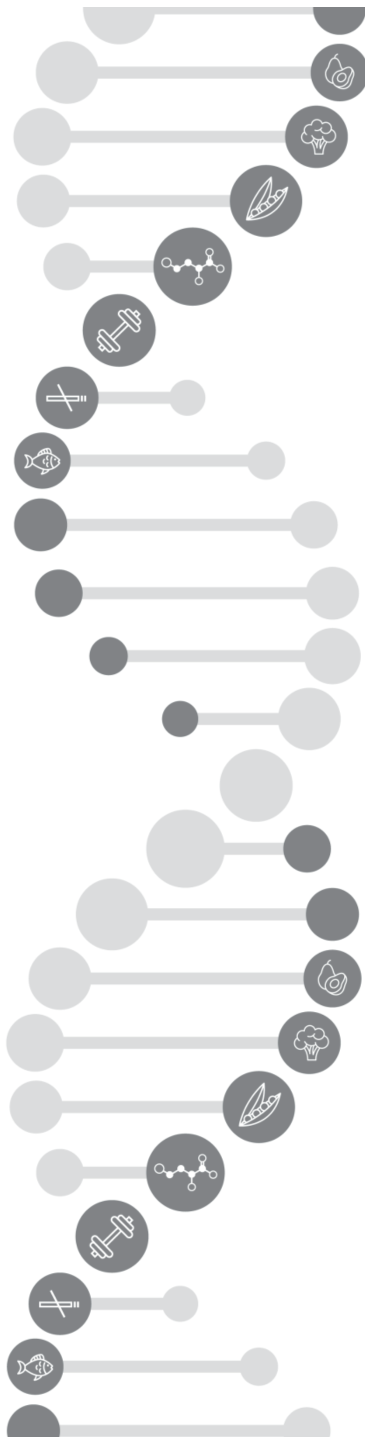
About the author 299

Words of gratitude 301

LIST OF ABBREVIATIONS

A	Adenine
ARIC	Atherosclerosis in Communities Study
BBMRI	Biobanking and BioMolecular resources Research Infrastructure
BHMT	Betaine-homocysteine methyltransferase
BIOS	Biobank-based Integrative Omics Studies Consortium
BMI	Body mass index
C	Cytosine
CBS	Cystathionine β-synthase
$-CH_3$	Methyl
CHARGE	Cohorts for Heart and Aging Research in Genomic Epidemiology
CHS	Cardiovascular Health Study
CODAM	Cohort on Diabetes and Atherosclerosis Maastricht
CpG	Cytosine-phosphate-Guanine
CSE	Cystathionine γ-lyase
DHFR	Dihydrofolate reductase
DMP	Differentially methylated CpG position
DMR	Differentially methylated CpG region
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase enzymes
EGCUT	Estonian Genome Center, University of Tartu
EPIC	European Prospective Investigation into Cancer
EPIC- Norfolk	European Prospective Investigation into Cancer and Nutrition-Norfolk
ERGO	Erasmus Rotterdam Gezondheid Onderzoek
EWAS	Epigenome-wide association study
FHS	Framingham Heart Study
F5L	French-Canadian family study on Factor V Leiden
G	Guanine
GENOA	Genetic Epidemiology Network of Arteriopathy
GOLDN	Genetics of Lipid Lowering Drugs and Diet Network
GRS	Genetic risk score
GTP	Grady Trauma Project
GWAS	Genome-wide association studies
Hcy	Homocysteine
HHcy	Hyperhomocysteinemia
5hmC	5-hydroxymethylcytosine

HPLC	High performance liquid chromatography
InCHIANTI	Invecchiare in Chianti
KORA	Cooperative Health Research in the Region of Augsburg
LBC	Lothian Birth Cohort
LL	LifeLines
LLD	LifeLinesDeep
LLS	Leiden Longevity Study
LC-MS/MS	HPLC tandem mass spectrometry
LUMA	Luminometric methylation assay
MARTHA	MARseille THrombosis Association Study
MAT	Methionine adenosyltransferase
MDB	Methyl-CpG-binding domain
MDBP	Methylated DNA-binding proteins
MeDIP	Methylated DNA immunoprecipitation
MESA	Multi Ethnic Study of Atherosclerosis
MeQTLs	Methylation Quantitative Trait Loci
MR	Mendelian Randomization
MSR	Methionine synthase reductase
MTHFR	5,10-Methylenetetrahydrofolate reductase
MTR/MS	Methionine synthase
NAS	Normative Aging Study
NK	Natural killer
NTR	Netherlands Twins Register
PAN	Prospective ALS Study Netherlands
RS	Rotterdam Study
SAH	S-adenosylhomocysteine
SAM	S-adenosylmethionine
SAHH	S-adenyl-l-homocysteine hydrolase
SHMT	Serine hydroxymethyltransferase
SNP	Single Nucleotide Polymorphism
T	Thymidine
THF	Tetrahydrofolate
WBC	Whole blood cell
YFS	Young Finns Study



PART A

Introduction

CHAPTER 1

General introduction and outline of thesis

Epigenetics is defined as the study of heritable changes in gene expression resulting from changes in a chromosome without alterations in its DNA sequence [1-3]. These epigenetic changes or chemical modifications occur either to the DNA, DNA binding histones or histone binding nucleosomes, and are referred to as DNA methylation, histone modifications and nucleosome positioning, respectively [4]. In normal processes, epigenetic modifications are important in regulating gene and non-coding RNA expression in response to changing conditions and maintaining normal development. DNA methylation is currently the most widely studied epigenetic mechanism.

DNA methylation is the addition of a methyl group to DNA and typically occurs at the 5' carbon position of the cytosine base in DNA to form 5-methylcytosine [Figure 1]. This process is catalyzed by DNA methyltransferase enzymes (DNMTs) [5] and frequently occurs at a Cytosine-phosphate-Guanine (CpG) dinucleotide where the cytosine base is followed by a guanine base separated by a phosphodiester bond. There are three major types of DNMTs: DNMT1, DNMT3a and DNMT3b. In mammalian cells, DNMT1 is most abundant of the three DNMTs [6]. It is referred to as maintenance methyltransferase because it prefers to bind to CpG sites of the hemi-methylated DNA (when only one of two complementary strands are methylated) in order to copy the existing methylation pattern of the DNA [7]. DNMT3a and DNMT3b are referred to as *de novo* methyltransferases because they are necessary for early development. They can bind to both hemi-methylated and non-methylated DNA [7].

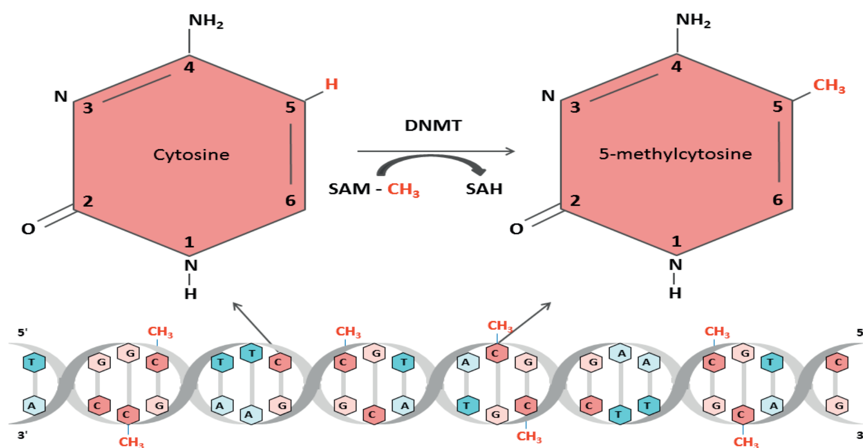


Figure 1. Methylation of cytosine to 5-methylcytosine, catalyzed by DNMT enzymes. The methyl transfer s-adenosylmethionine (SAM) donates its methyl group to cytosine to produce s-adenosylhomocysteine (SAH).

DNA methylation can influence gene function and gene expression by two mechanisms. One, CpG methylation can interfere with the binding of the transcription factors to the DNA, which can in turn suppress gene transcription. Two, there are certain proteins that act as transcriptional repressors by binding to specific sequences of the methylated DNA and are called methylated DNA-binding proteins (MDBP) [8-10]. DNA methylation is dynamic across the life course [11] and age-associated changes have been seen both during early childhood [12] and adulthood [13, 14]. Sex-associated DNA methylation differences have also been found [15]. DNA methylation is an essential epigenetic process involved in regulation of genes in all biological processes and has an essential role in embryonic development, genomic imprinting, genomic integrity, transcriptional regulation and adult homeostasis [16, 17]. DNA methylation is influenced by several metabolic, environmental, nutritional, lifestyle and genetic factors [Figure 2].

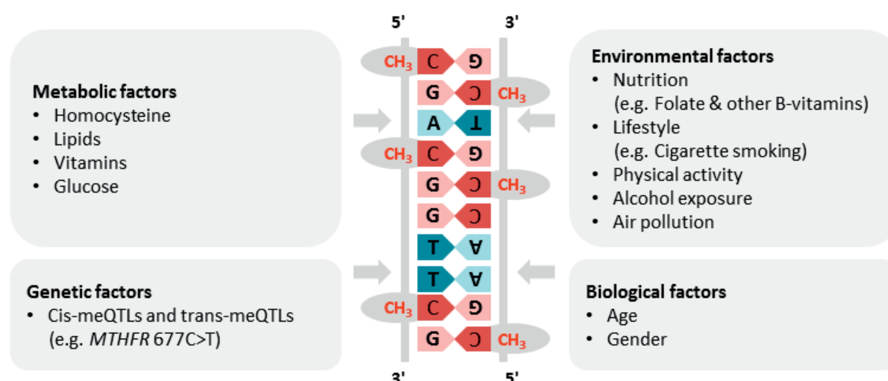


Figure 2. Factors that can affect DNA methylation. Nucleotides in blue are the Adenine (A) and Thymidine (T) bases and nucleotides in purple are the Guanine (G) and Cytosine (C) bases, respectively. Methyl groups in red typically attach to the 5' carbon position of the cytosine base of DNA to form 5-methylcytosine. Cis-meQTLs (methylation Quantitative Trait Loci) are the SNPs that correlate with a CpG that is <1 Mb away. Trans-meQTLs are SNPs that correlate with CpGs that are >1 Mb away.

METHYL-GROUP METABOLISM AND ROLE OF B-VITAMINS

Methyl-group metabolism, also known as one-carbon metabolism, is a network of biochemical reactions that is involved in the transfer of one-carbon units, typically as methyl ($-CH_3$) groups, from one metabolite to the other, and subsequently to the DNA [18]. Several enzymes participate in the catalysis of these reactions along with dietary micronutrients that act as cofactors, such as B-vitamins (folate, vitamin B2, B6 and B12), choline and betaine [19] [Figure 3]. For this reason, nutrition status, particularly micronutrient intake, has been a focal point when investigating epigenetic mechanisms

[18]. The important metabolites of this pathway include s-adenosylmethionine (SAM), s-adenosylhomocysteine (SAH) and homocysteine (Hcy). In the transmethylation pathway, the methyl donor, SAM is biosynthesized from methionine and donates its methyl group to the DNA, itself being converted to SAH. SAH is a potent inhibitor of this same methyltransferase reaction and is hydrolyzed to Hcy. Due to the reversible reaction of Hcy with SAH, Hcy needs to be transported out of the cell or be metabolized to prevent accumulation of SAH. Hcy metabolism takes place either through the remethylation pathway by converting back to methionine, or the transsulfuration pathway by converting to cystathionine [20]. Impairments in the metabolism of Hcy can lead to elevated Hcy and consequently elevated SAH and reduced SAM:SAH ratio [21, 22]. Elevated Hcy has been associated with reduced global methylation [22] and is a marker of several disorders such as cardiovascular diseases, neural tube defects, cognitive decline and different types of cancers [23]. Elevated Hcy can be a consequence of either genetic variants in one or more of the key enzymes and/or nutritional deficiencies [24].

The most widely studied Hcy-associated genetic polymorphism includes the methylene tetrahydrofolate reductase (*MTHFR*) 677C>T variant. This variant can lead to reduced MTHFR enzyme activity of up to 45% and impairs the conversion of 5,10-methylenetetrahydrofolate to 5-methyltetrahydrofolate (5-methyl-THF) in the remethylation pathway, where 5-MTHF acts as a cosubstrate by converting Hcy to methionine [24, 25]. Another commonly studied polymorphism for MTHFR includes *MTHFR* 1298A>C, which is associated with 68% of reduced MTHFR activity [25, 26]. In the remethylation pathway, methionine synthase (MTR) catalyzes the remethylation of Hcy to methionine. The common polymorphism within the MTR gene is *MTR* A66G, which results in reduced activity of the MTR enzyme and Hcy accumulation [27]. In the transsulfuration pathway, cystathionine-beta-synthase (CBS) enzyme catalyses the conversion of Hcy to cystathionine. Two mutations in the CBS enzyme include the rare *CBS* 1330G>A [28] and *CBS* 833T>C [29]. Both these mutations are associated with reduced activity of CBS enzyme and Hcy accumulation and occur in CBS deficiency. Apart from the ones mentioned, other Hcy-associated polymorphisms were also previously found in candidate gene studies and genome-wide association studies (GWAS) [25, 30-32]. In a previous large GWAS of 44,147 individuals, van Meurs JB et al found 18 variants associated with Hcy, of which 13 were common polymorphisms that explained 5.9% of variation in Hcy [32].

Folate and vitamin B12 are widely studied co-factors, deficiencies of which could impair the methyl-group metabolism and lead to elevated Hcy [24]. **Therefore, impaired methyl-group metabolism due to shortage of B-vitamins or genetic variants could play a vital role in DNA methylation.** A few studies until now have investigated the association of Hcy, Hcy-associated variants particularly the *MTHFR* 677C>T and B-vitamins with global DNA methylation, but showed inconsistent findings.

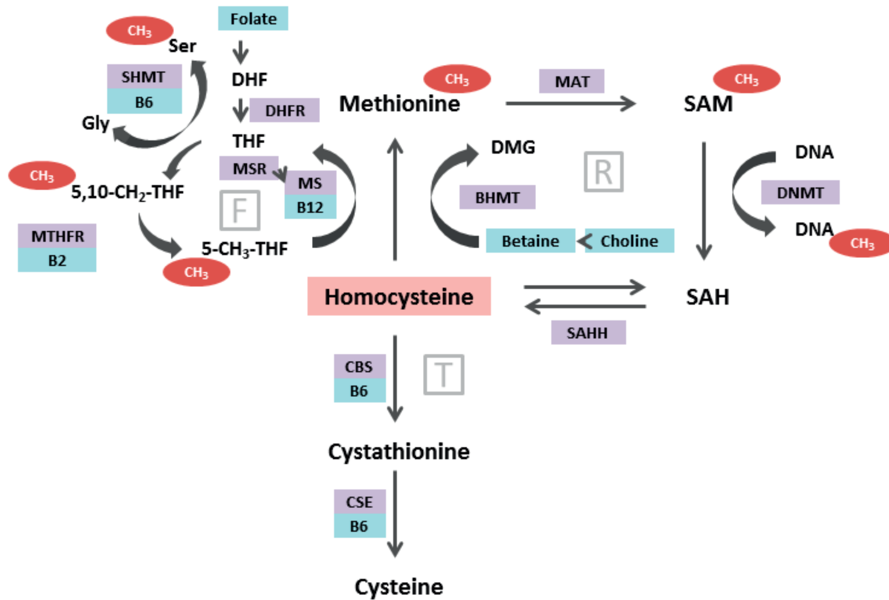


Figure 3. A simplified figure of methyl-group metabolism, which includes the main enzymes (purple) and dietary factors (blue). DHFR: Dihydrofolate reductase; SHMT: Serine hydroxymethyltransferase; MTHFR: 5,10-Methylenetetrahydrofolate reductase; MSR: Methionine synthase reductase; MS: Methionine synthase; BHMT: Betaine-homocysteine methyltransferase; MAT: Methionine adenosyltransferase; DNMT: DNA methyltransferase; SAHH: S-adenyl-l-homocysteine hydrolase; CBS: Cystathionine β-synthase; CSE: Cystathionine γ-lyase. Ser: Serine; Gly: Glycine. R: Remethylation pathway; T: Transsulfuration pathway; F: Folate cycle.

LIFESTYLE FACTORS

In addition to the one carbon metabolism and nutritional factors, other lifestyle factors such as cigarette smoking, alcohol intake and physical activity can also affect DNA methylation [33]. Cigarette smoking is a crucial risk factor for disorders such as respiratory diseases, cardiovascular diseases, cancer and reproductive outcomes [34]. There are a few mechanisms known by which cigarette smoking can alter DNA methylation [35]. One, cigarette smoke contains carcinogens such as arsenic, chromium, polycyclic aromatic hydrocarbons, formaldehyde, and nitrosamines that causes double-stranded breaks to the DNA causing DNA damage [36]. The enzyme DNMT1 is recruited to repair the damage in survival cells which causes methylation of CpGs adjacent to the repaired nucleotides [36-38]. Two, cigarette smoke can alter DNA methylation through the effect of *nicotine* on DNMT1 expression. Nicotine binds to and activates the nicotinic acetylcholine receptors that increases intracellular calcium and

activates the key transcription factor (cAMP response element-binding protein) of many genes with possible downregulation of DNMT1 [39-41]. Three, cigarette smoking increases expression of the DNA-binding protein, Sp1, which protects CpG sites from being methylated [42-44]. Four, cigarette smoke contains carbon monoxide that competes with oxygen to bind to hemoglobin. This causes inadequate oxygenation of tissues in a condition called hypoxia [45]. Hypoxia further leads to the HIF-1α-dependent upregulation of methionine adenosyltransferase 2A, which is an enzyme that synthesizes the methyl donor, SAM [46]. Previous EWAS studies have investigated the association between cigarette or tobacco smoking and epigenome-wide DNA methylation [47-56], but large-scale meta-analysis studies have not been conducted yet. Sufficiently large sample size is necessary for detecting small effects, reliability of results and internal and external validation [57, 58]. Such studies may further help in identifying new smoking-related CpGs that can serve as biomarkers in smoking-related pathologies.

METHODS TO MEASURE DNA METHYLATION

Different methods that are used to profile global and site-specific genome-wide DNA methylation are available [59-64]. These methods vary in terms of their genomic region coverage, DNA input and resolution. For global methylation, the high performance liquid chromatography (HPLC) and HPLC tandem mass spectrometry (LC-MS/MS) are regarded as the golden standards to measure the total methylated cytosine content in the DNA [60, 61]. These methods are highly quantitative and reproducible. Since the protocol for their assay optimization is demanding and requires relatively large amounts of DNA (1-5 µg), other methods have been developed that use more readily available equipment, require less DNA and are less expensive. These include PCR-based methods which measure methylation status of genomic repeat elements such as LINE-1 and Alu, and other method includes the Luminometric methylation assay (LUMA) which is based on polymerase extension assay using the pyrosequencing platform [60, 61]. The repeat elements are representative of the human genome since they constitute about 45% of it [60]. However, the information on methylation levels derived from these surrogate methods is limited to analyzed sequences and their comparison to total methyl cytosine content in DNA is uncertain. LINE-1 assay showed an acceptable surrogate to golden standard methods as compared to Alu assay or LUMA [60, 61]. Considering the highest sensitivity and specificity, minimal assay-to-assay variability and also the amount of starting material required by each assay, these methods are recommended in order of their preference: (1) LINE-1/pyrosequencing; (2) LC-MS/MS; and (3) LUMA [61]. For gene-specific methods, PCR-based methods were developed wherein primers are used to amplify the gene of interest from the bisulfite converted DNA [62].

For genome-wide site-specific DNA methylation profiling, various microarray- and sequencing-based technologies have been developed [59, 61, 62, 64]. They are grouped into 3 categories, namely; (1) restriction enzyme based methods, (2) affinity enrichment-based methods, and (3) bisulfite conversion-based methods. Restriction enzyme based methods cleave DNA at specific positions to distinguish between methylated and unmethylated DNA. Affinity enrichment-based methods use either anti-methylcytosine antibodies (as in methylated DNA immunoprecipitation (MeDIP)) or methyl-CpG-binding domain (MDB) proteins specific for 5-methyl cytosines to enrich methylated DNA regions. Bisulphite conversion-based methods can distinguish cytosine base from methylated cytosine by converting cytosines to uracil and keeping methylated cytosines as cytosines. These three types of methods are either coupled with microarray or sequencing technologies. Each of these techniques have their own pros and cons, and the preference for a particular technique highly depends on the research questions, the amount of required starting material, sensitivity, specificity, resolution, coverage and cost. Comparing with other techniques, bisulphite conversion-based Illumina HumanMethylation450K BeadChip arrays (Illumina, Inc., San Diego, CA, USA) provide a high coverage with high sensitivity and specificity at a single base resolution, along with lower cost and low starting material. This makes it as the preferred technique for our study.

EPIGENOME-WIDE ASSOCIATION STUDIES

In order to investigate associations between site-specific DNA methylation variations and a phenotype of interest, EWASs assess a genome-wide collection of epigenetic marks that exist in a cell at any given point in time [65]. Illumina 450K array measures up to 485,764 CpGs per DNA sample distributed across the whole genome [66]. These CpGs represent 99% of RefSeq genes and 96% of CpG island regions that include high density islands and low density shores and shelves. This array also includes non-CpG methylated sites identified in human stem cells (CHH sites), differentially methylated CpG positions (DMPs) identified in tumor versus normal, FANTOM 4 promoters, DNase hypersensitive sites and miRNA promoter regions. Validation and evaluation of this technique using cell lines and tissue samples indicated that it is highly robust and reliable and showed high concordance with the widely used pyrosequencing method [66-68].

The human genome contains more than 28 million CpGs [69], 70-80% of which are methylated [70, 71]. The Illumina 450K arrays measures 1.7% of the CpGs in the genome. Among the gene regions for CpGs included on this array, most (41%) of them are present at gene promoters, 31% at gene bodies, 3% at 3'UTR regions and 25% at intergenic regions. Among the genome regions, most (31%) of them are present in high density at regions known as "CpG islands" that are usually associated with promoters. CpGs with low density (23%) are present at "CpG island shores" that are

about 2 Kb away from CpG islands. CpGs with much lower density (31%) than CpG island shores are present at “CpG island shelves” also known as CpG poor regions, that are in turn about 2 Kb away from CpG island shores. The rest (36%) of them are the isolated CpGs known as “Open Sea” regions [66] [Figure 4]. This gives an advantage over other techniques that measure global methylation as the total methylated cytosine content of the DNA, which does not indicate the specific CpGs, genes and pathways involved. Illumina 450k arrays give opportunities to uncover novel epigenetic markers and help explore mechanisms that could be associated with disease risks in a hypothesis-free manner. When the knowledge of associations about a particular phenotype and DNA methylation is limited, the hypothesis free approach gives the benefit to eliminate the preexisting bias of traditional biology, meaning that it is unbiased by prior pathophysiological assumptions [72]. Therefore, hypothesis-free approach gives us the opportunity to unravel novel epigenetic markers associated with methyl-group metabolism factors.

The design of the Illumina 450k arrays is such that it contains two different probes; type I (28% CpGs) and type II (72% CpGs) [73]. Type I probes use two different bead types corresponding to methylated and unmethylated probe alleles, both of whose signals are generated in same red colour channel. Type II assay uses one bead type corresponding to both methylated and unmethylated probe alleles whose signals are generated in green and red colour channels, respectively.

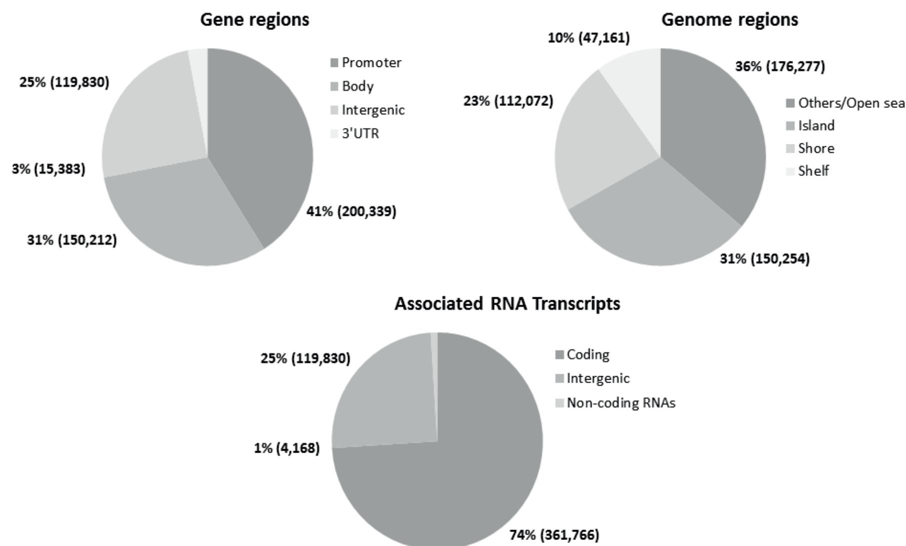


Figure 4. Distribution of CpGs in the Illumina 450k array based on gene regions, genomic regions and associated RNA transcripts. Figure adapted from Sandoval J et al.

DIFFERENTIALLY METHYLATED REGIONS

It is known that CpGs that are closely located to each other often show similar methylation patterns [74] and that DNA methylation is regulated in longer regions [75]. Therefore, in linear models, treating a region as a unit with two or more CpG sites might help identify regions with consistent methylation changes. Different methods have been developed that use different approaches to identify differentially methylated regions (DMRs) such as IMA [76], Bumphunter [77], DMRcate [78], Comb-p [79, 80], Probe Lasso [81], Aclust [82] and seqlm [83]. Authors of the DMRcate and seqlm compared their method with other methods and showed that their method performed better than the other methods [78, 83]. However, despite so many developed methods and their approaches, they all have their pros and cons, and there is no standard method until now. IMA package in R calculates methylation mean or median of β-values for gene-based predefined regions such as promoter, 5'-UTR, first exon, gene body and 3'-UTR and also other regions that are not necessarily gene-based, such as CpG islands, shores and shelves [76]. On the other hand, methods such as Bumphunter and DMRcate identify DMRs using region finding algorithms that are based on sliding window and effect size cutoffs [77, 78]. These methods however might not be well suited for the uneven distribution of the Illumina 450k CpGs. In the analysis from this thesis, we used the method Comb-p [79, 80] that was developed considering this uneven distribution. It uses nominal p-values as input in a sliding window and also takes into account the correlation between CpGs associated with these p-values.

MENDELIAN RANDOMIZATION STUDIES

Classical epidemiological association studies do not indicate the cause and effect of a particular association. They just detect association. To determine causal direction, genetic instrumental variables of phenotypes or methylation can be used, in a concept called Mendelian Randomization (MR) [84, 85]. The relationship of Hcy, vitamin B12 and folate with DNA methylation can be subject to substantial bias, given the strong relationship between several lifestyle factors, diseases and methyl-group metabolism factors. To circumvent this bias, genetic factors determining the methyl-group metabolism factors as an instrument can be used to study the relationship between these factors and methylation by MR. This would eliminate the effects that are possibly caused by measurement errors, confounding and reverse causality.

STUDY POPULATION

1. Rotterdam Study (RS)

All studies described in this thesis were performed within a large population-based cohort study of the Netherlands, the RS, also known as "Erasmus Rotterdam Gezondheid Onderzoek (ERGO)". The RS is a prospective study aimed at assessing the occurrence of risk factors for chronic (cardiovascular, endocrine, locomotor, hepatic,

neurological, ophthalmic, psychiatric, dermatological, oncological, and respiratory) diseases in the elderly [86]. The study comprises 14,926 subjects in total, living in the well-defined Ommoord district in the city of Rotterdam in the Netherlands. Genome-wide DNA-methylation levels were determined in a random subset of 1,613 individuals using the Illumina HumanMethylation450K BeadChip arrays [Figure 5].

Several EWASs of phenotype measures such as Hcy, the common Hcy associated variant MTHFR 677C>T, folate intake, vitamin B12 intake, cigarette smoking, age, gender, lipids, and anthropometry were conducted in RS. Inorder to improve the estimates of found associations and find new loci [57, 58], the power of the study was increased by including multiple studies in a meta-analysis approach. Several studies participated in the following consortium efforts to achieve the goal.

2. Biobank-based Integrative Omics Studies Consortium (BIOS) consortium

Within the BIOS consortium, Biobanking and BioMolecular resources Research Infrastructure (BBMRI)-Omics (http://www.bbmri.nl/) is the joint collection of omics data that has collaboratively been generated and is made available for BBMRI researchers focusing on integrative omics studies in Dutch Biobanks in the Netherlands. RNA-sequencing (>15 M paired end reads) and genome-wide DNA methylation data has been generated using Illumina 450k arrays and Illumina RNA sequencing, for over 4000 samples. In addition, phenotype measures such as age, smoking, gender, lipids, metabolomics (>200 metabolites) and anthropometry are available within the same samples. The Netherlands-based cohorts that are a part of this consortium are the Rotterdam Study, the Leiden Longevity Study, the Groningen LifeLines study, the Netherlands Twins Register, the Cohort on Diabetes and Atherosclerosis Maastricht and the Prospective ALS Study Netherlands [Figure 6].

3. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium

The CHARGE consortium (http://www.chargeconsortium.com/) was formed to facilitate genome-wide association study meta-analyses and replication opportunities among multiple large and well-phenotyped longitudinal cohort studies [87]. This consortium is a collaboration between many cohorts studies, both from Europe and the United States. The five founding member cohorts of this effort include the Age, Gene/Environment Susceptibility-Reykjavik Study, the Atherosclerosis Risk in Communities Study, the Cardiovascular Health Study, the Framingham Heart Study and the Rotterdam Study. The additional core cohorts include the Coronary Artery Risk Development in Young Adults, the Family Heart Study, the Health, Aging, and Body Composition Study, the Jackson Heart Study and the Multi-Ethnic Study of Atherosclerosis. Within this thesis, cohorts within this consortium who had both DNA methylation data measured from the Illumina 450k arrays as well as the phenotypes of interest, were used for the EWAS analyses [Figure 6].

OUTLINE OF THIS THESIS

The aim of this thesis is to study the association of methyl-group metabolism, nutritional and lifestyle factors with genome-wide DNA methylation. The first section of this thesis includes two literature reviews. **Chapter 2** is a literature review of Hcy and its role in DNA methylation. **Chapter 3** is a systematic literature review of the relation between micro- and macro- nutrients and DNA methylation in humans across the life course.

The second section of this thesis focuses on a key metabolite of the methyl-group metabolism, Hcy. **Chapter 4** is a meta-analysis of EWASs to investigate the association between plasma Hcy and DNA methylation in leukocytes of 2,035 individuals from six cohorts. In **Chapter 5**, we used genetically defined elevated Hcy as an instrument, i.e. the *MTHFR* 677C>T variant and the combined weighted genetic risk score of 18 previously studied Hcy-associated variants, to test whether genetically defined elevated Hcy levels are associated with DNA methylation changes in leukocytes of 9,894 individuals from 12 cohorts. In **Chapter 6**, we conducted an interaction study to investigate the effect of elevated Hcy in individuals by *MTHFR* 677C>T genotype on genome-wide DNA methylation in leukocytes of 1280 individuals from 2 cohorts.

The third section of this thesis focuses on nutrition and lifestyle factors. **Chapter 7** is a meta-analysis of EWASs to investigate the association of folate intake and vitamin B12 intake with DNA methylation in leukocytes of 5,841 participants from 10 cohorts. **Chapter 8** focuses on association between cigarette smoking as a lifestyle factor and DNA methylation in leukocyte assessed in 15,907 individuals (2,433 current, 6,518 former, and 6,956 never smokers) from 16 cohorts.

Figure 5. Overview of the long-term population survey of the Rotterdam Study (RS), also known as “Erasmus Rotterdam Gezondheid Onderzoek” (ERGO) in the Netherlands. In 1989, the first cohort, RS-I (also known as ERGO) comprised of 7,983 subjects with age 55 years or above. In 2000, the second cohort, RS-II (also known as ERGOPLUS) was included with 3,011 subjects who had reached an age of 55 or over. In 2006, the third cohort, RS-III (also known as EROGJONG) was further included with 3,932 subjects with age 45 years and above. In 2016, the forth cohort, RS-IV (also known as ERGOXTRA) was included with expected recruitment of 4,000 subjects with age 40 years and above. At the Genetic Laboratory (Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, the Netherlands), genome-wide DNA-methylation levels in 1,613 individuals mainly belonging to ERGOJONG and ERGO-5 were determined using the Illumina HumanMethylation450K BeadChip arrays (Illumina, Inc., San Diego, CA, USA). After quality control, 1,544 of the 1,613 individuals were used for the downstream analysis. Phenotypes of interest such as homocysteine, MTHFR 677C>T, B-vitamin intake and cigarette smoking with overlapping DNA methylation data were used to run EWASs in this thesis.

RS-I cohort
RS-II cohort
RS-III cohort
RS-IV cohort
Project name
DNA methylation N = 1,544
Homocysteine N = 700
MTHFR 677C>T N = 731
Dietary B-vitamins N = 900
Cigarette smoking N = 686
RS-I-1
RS-I-2
RS-I-3
RS-I-4
RS-I-5
RS-I-6
RS-I-7
RS-II-1
RS-II-2
RS-II-3
RS-II-4
RS-II-5
RS-III-1
RS-III-2
RS-III-3
RS-IV-1
ERGO-1
ERGO-2
ERGO-3
ERGOPLUS
ERGO-4
ERGOJONG
ERGO-5
ERGO-6
ERGOXTRA
ERGO-7
n = 1
n = 1
n = 32
n = 731
n = 700
n = 731
n = 417
n = 686
n = 779
n = 483
n = 731

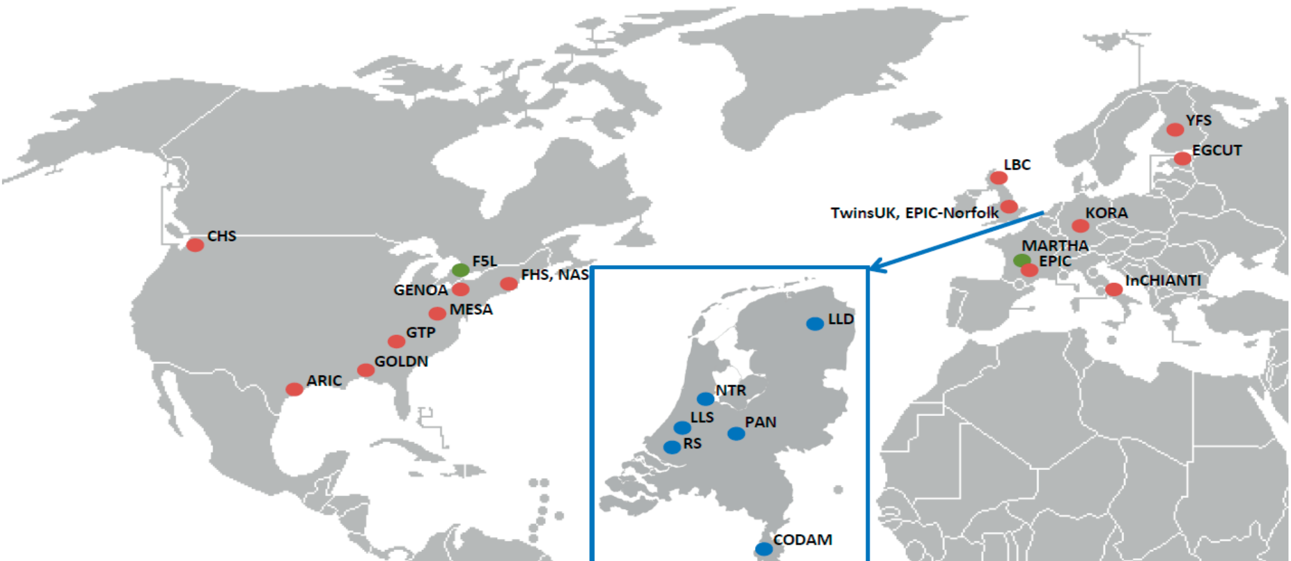


Figure 6. Overview of the collaborating cohorts in this thesis. (1) BIOS consortium (Blue): RS (Rotterdam Study), LLS (Leiden Longevity Study), LLD (Groningen LifeLines study), NTR (Netherlands Twins Register), CODAM (Cohort on Diabetes and Atherosclerosis Maastricht) and PAN (Prospective ALS Study Netherlands). (2) CHARGE consortium (Red): RS (Rotterdam Study), LBC (Lothian Birth Cohort), Cooperative Health Research in the Region of Augsburg (KORA), TwinsUK, Invecchiare in Chianti (InCHIANTI) study, Framingham Heart Study (FHS), Genetics of Lipid Lowering Drugs and Diet Network (GOLDN), Atherosclerosis in Communities Study (ARIC), Cardiovascular Health Study (CHS), Multi Ethnic Study of Atherosclerosis (MESA), Grady Trauma Project (GTP), European Prospective Investigation into Cancer (EPIC), European Prospective Investigation into Cancer and Nutrition-Norfolk (EPIC-Norfolk), Genetic Epidemiology Network of Arteriopathy (GENOA), Estonian Genome Center, University of Tartu (EGCUT), Young Finns Study (YFS) and Normative Aging Study (NAS) (3) Other studies (Green): MARseille THrombosis Association Study (MARTHA) and French-Canadian family study on Factor V Leiden (F5L) thrombophilia.

REFERENCES

1. Berger SL, Kouzarides T, Shiekhattar R, Shilatifard A. An operational definition of epigenetics. Genes Dev 2009;23:781-3.
2. Goldberg AD, Allis CD, Bernstein E. Epigenetics: a landscape takes shape. Cell 2007;128:635-8.
3. Bird A. Perceptions of epigenetics. Nature 2007;447:396-8.
4. Portela A, Esteller M. Epigenetic modifications and human disease. Nat Biotechnol 2010;28:1057-68.
5. Bestor TH. Cloning of a mammalian DNA methyltransferase. Gene 1988;74:9-12.
6. Robertson KD, Uzvolgyi E, Liang G, Talmadge C, Sumegi J, Gonzales FA, Jones PA. The human DNA methyltransferases (DNMTs) 1, 3a and 3b: coordinate mRNA expression in normal tissues and overexpression in tumors. Nucleic Acids Res 1999;27:2291-8.
7. Hermann A, Goyal R, Jeltsch A. The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. J Biol Chem 2004;279:48350-9.
8. Jones PA, Laird PW. Cancer epigenetics comes of age. Nat Genet 1999;21:163-7.
9. Kass SU, Pruss D, Wolffe AP. How does DNA methylation repress transcription? Trends Genet 1997;13:444-9.
10. Tate PH, Bird AP. Effects of DNA methylation on DNA-binding proteins and gene expression. Curr Opin Genet Dev 1993;3:226-31.
11. Heyn H, Li N, Ferreira HJ, Moran S, Pisano DG, Gomez A, et al. Distinct DNA methylomes of newborns and centenarians. Proc Natl Acad Sci U S A 2012;109:10522-7.
12. Xu CJ, Bonder MJ, Soderhall C, Bustamante M, Baiz N, Gehring U, et al. The emerging landscape of dynamic DNA methylation in early childhood. BMC Genomics 2017;18:25.
13. Kananen L, Marttila S, Nevalainen T, Jylhava J, Mononen N, Kahonen M, et al. Aging-associated DNA methylation changes in middle-aged individuals: the Young Finns study. BMC Genomics 2016;17:103.
14. Reynolds LM, Taylor JR, Ding J, Lohman K, Johnson C, Siscovick D, et al. Age-related variations in the methylome associated with gene expression in human monocytes and T cells. Nat Commun 2014;5:5366.
15. Inoshita M, Numata S, Tajima A, Kinoshita M, Umehara H, Yamamori H, et al. Sex differences of leukocytes DNA methylation adjusted for estimated cellular proportions. Biol Sex Differ 2015;6:11.
16. Li E, Beard C, Jaenisch R. Role for DNA methylation in genomic imprinting. Nature 1993;366:362-5.
17. Messerschmidt DM, Knowles BB, Solter D. DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. Genes Dev 2014;28:812-28.
18. Anderson OS, Sant KE, Dolinoy DC. Nutrition and epigenetics: an interplay of dietary methyl donors, one-carbon metabolism and DNA methylation. J Nutr Biochem 2012;23:853-9.
19. Selhub J. Folate, vitamin B12 and vitamin B6 and one carbon metabolism. J Nutr Health Aging 2002;6:39-42.
20. Blom HJ, Smulders Y. Overview of homocysteine and folate metabolism. With special references to cardiovascular disease and neural tube defects. J Inherit Metab Dis 2011;34:75-81.
21. Fu W, Dudman NP, Perry MA, Young K, Wang XL. Interrelations between plasma homocysteine and intracellular S-adenosylhomocysteine. Biochem Biophys Res Commun 2000;271:47-53.

22. James SJ, Melnyk S, Pogribna M, Pogribny IP, Caudill MA. Elevation in S-adenosylhomocysteine and DNA hypomethylation: potential epigenetic mechanism for homocysteine-related pathology. J Nutr 2002;132:2361S-2366S.
23. Stover PJ. One-carbon metabolism-genome interactions in folate-associated pathologies. J Nutr 2009;139:2402-5.
24. Selhub J. Homocysteine metabolism. Annu Rev Nutr 1999;19:217-46.
25. Sunder-Plassmann G, Fodinger M. Genetic determinants of the homocysteine level. Kidney Int Suppl 2003:S141-4.
26. Weisberg I, Tran P, Christensen B, Sibani S, Rozen R. A second genetic polymorphism in methylenetetrahydrofolate reductase (MTHFR) associated with decreased enzyme activity. Mol Genet Metab 1998;64:169-72.
27. Leclerc D, Wilson A, Dumas R, Gafuik C, Song D, Watkins D, et al. Cloning and mapping of a cDNA for methionine synthase reductase, a flavoprotein defective in patients with homocystinuria. Proc Natl Acad Sci U S A 1998;95:3059-64.
28. Kluijtmans LA, Boers GH, Stevens EM, Renier WO, Kraus JP, Trijbels FJ, et al. Defective cystathionine beta-synthase regulation by S-adenosylmethionine in a partially pyridoxine responsive homocystinuria patient. J Clin Invest 1996;98:285-9.
29. Masud R, Qureshi IZ. Tetra primer ARMS-PCR relates folate/homocysteine pathway genes and ACE gene polymorphism with coronary artery disease. Mol Cell Biochem 2011;355:289-97.
30. Clifford AJ, Chen K, McWade L, Rincon G, Kim SH, Holstege DM, et al. Gender and single nucleotide polymorphisms in MTHFR, BHMT, SPTLC1, CRBP2, CETP, and SCARB1 are significant predictors of plasma homocysteine normalized by RBC folate in healthy adults. J Nutr 2012;142:1764-71.
31. Lange LA, Croteau-Chonka DC, Marvelle AF, Qin L, Gaulton KJ, Kuzawa CW, et al. Genome-wide association study of homocysteine levels in Filipinos provides evidence for CPS1 in women and a stronger MTHFR effect in young adults. Hum Mol Genet 2010;19:2050-8.
32. van Meurs JB, Pare G, Schwartz SM, Hazra A, Tanaka T, Vermeulen SH, et al. Common genetic loci influencing plasma homocysteine concentrations and their effect on risk of coronary artery disease. Am J Clin Nutr 2013;98:668-76.
33. Lim U, Song MA. Dietary and lifestyle factors of DNA methylation. Methods Mol Biol 2012;863:359-76.
34. National Center for Chronic Disease P, Health Promotion Office on S, Health. 2014.
35. Lee KW, Pausova Z. Cigarette smoking and DNA methylation. Front Genet 2013;4:132.
36. Huang J, Okuka M, Lu W, Tsibris JC, McLean MP, Keefe DL, Liu L. Telomere shortening and DNA damage of embryonic stem cells induced by cigarette smoke. Reprod Toxicol 2013;35:89-95.
37. Mortusewicz O, Schermelleh L, Walter J, Cardoso MC, Leonhardt H. Recruitment of DNA methyltransferase I to DNA repair sites. Proc Natl Acad Sci U S A 2005;102:8905-9.
38. Cuozzo C, Porcellini A, Angrisano T, Morano A, Lee B, Di Pardo A, et al. DNA damage, homology-directed repair, and DNA methylation. PLoS Genet 2007;3:e110.
39. Lee EW, D'Alonzo GE. Cigarette smoking, nicotine addiction, and its pharmacologic treatment. Arch Intern Med 1993;153:34-48.
40. Shen JX, Yakel JL. Nicotinic acetylcholine receptor-mediated calcium signaling in the nervous system. Acta Pharmacol Sin 2009;30:673-80.
41. Satta R, Maloku E, Zhubi A, Pibiri F, Hajos M, Costa E, Guidotti A. Nicotine decreases DNA methyltransferase 1 expression and glutamic acid decarboxylase 67 promoter methylation in GABAergic interneurons. Proc Natl Acad Sci U S A 2008;105:16356-61.

42. Mercer BA, Wallace AM, Brinckerhoff CE, D'Armiento JM. Identification of a cigarette smoke-responsive region in the distal MMP-1 promoter. Am J Respir Cell Mol Biol 2009;40:4-12.
43. Di YP, Zhao J, Harper R. Cigarette smoke induces MUC5AC protein expression through the activation of Sp1. J Biol Chem 2012;287:27948-58.
44. Kadonaga JT, Carner KR, Masiarz FR, Tjian R. Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. Cell 1987;51:1079-90.
45. Olson KR. Carbon monoxide poisoning: mechanisms, presentation, and controversies in management. J Emerg Med 1984;1:233-43.
46. Liu Q, Liu L, Zhao Y, Zhang J, Wang D, Chen J, et al. Hypoxia induces genomic DNA demethylation through the activation of HIF-1alpha and transcriptional upregulation of MAT2A in hepatoma cells. Mol Cancer Ther 2011;10:1113-23.
47. Breitling LP, Yang R, Korn B, Burwinkel B, Brenner H. Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. Am J Hum Genet 2011;88:450-7.
48. Breitling LP, Salzmann K, Rothenbacher D, Burwinkel B, Brenner H. Smoking, F2RL3 methylation, and prognosis in stable coronary heart disease. Eur Heart J 2012;33:2841-8.
49. Wan ES, Qiu W, Baccarelli A, Carey VJ, Bacherman H, Rennard SI, et al. Cigarette smoking behaviors and time since quitting are associated with differential DNA methylation across the human genome. Hum Mol Genet 2012;21:3073-82.
50. Wan ES, Qiu W, Carey VJ, Morrow J, Bacherman H, Foreman MG, et al. Smoking-Associated Site-Specific Differential Methylation in Buccal Mucosa in the COPDGene Study. Am J Respir Cell Mol Biol 2015;53:246-54.
51. Zeilinger S, Kuhnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, et al. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. PLoS One 2013;8:e63812.
52. Shenker NS, Ueland PM, Polidoro S, van Veldhoven K, Ricceri F, Brown R, et al. DNA methylation as a long-term biomarker of exposure to tobacco smoke. Epidemiology 2013;24:712-6.
53. Shenker NS, Polidoro S, van Veldhoven K, Sacerdote C, Ricceri F, Birrell MA, et al. Epigenome-wide association study in the European Prospective Investigation into Cancer and Nutrition (EPIC-Turin) identifies novel genetic loci associated with smoking. Hum Mol Genet 2013;22:843-51.
54. Guida F, Sandanger TM, Castagne R, Campanella G, Polidoro S, Palli D, et al. Dynamics of smoking-induced genome-wide methylation changes with time since smoking cessation. Hum Mol Genet 2015;24:2349-59.
55. Qiu W, Wan E, Morrow J, Cho MH, Crapo JD, Silverman EK, DeMeo DL. The impact of genetic variation and cigarette smoke on DNA methylation in current and former smokers from the COPDGene study. Epigenetics 2015;10:1064-73.
56. Gao X, Jia M, Zhang Y, Breitling LP, Brenner H. DNA methylation changes of whole blood cells in response to active smoking exposure in adults: a systematic review of DNA methylation studies. Clin Epigenetics 2015;7:113.
57. Michels KB, Binder AM, Dedeurwaerder S, Epstein CB, Greally JM, Gut I, et al. Recommendations for the design and analysis of epigenome-wide association studies. Nat Methods 2013;10:949-55.
58. Lin X, Barton S, Holbrook JD. How to make DNA methylome wide association studies more powerful. Epigenomics 2016;8:1117-29.
59. Yong WS, Hsu FM, Chen PY. Profiling genome-wide DNA methylation. Epigenetics Chromatin 2016;9:26.

60. Lisanti S, Omar WA, Tomaszewski B, De Prins S, Jacobs G, Koppen G, et al. Comparison of methods for quantification of global DNA methylation in human cells and tissues. PLoS One 2013;8:e79044.
61. Kurdyukov S, Bullock M. DNA Methylation Analysis: Choosing the Right Method. Biology (Basel) 2016;5.
62. Harrison A, Parle-McDermott A. DNA methylation: a timeline of methods and applications. Front Genet 2011;2:74.
63. Guerrero-Bosagna C. DNA Methylation Research Methods. Mater Methods 2013;3:206.
64. Bibikova M, Fan JB. Genome-wide DNA methylation profiling. Wiley Interdiscip Rev Syst Biol Med 2010;2:210-23.
65. Rakyan VK, Down TA, Balding DJ, Beck S. Epigenome-wide association studies for common human diseases. Nat Rev Genet 2011;12:529-41.
66. Sandoval J, Heyn H, Moran S, Serra-Musach J, Pujana MA, Bibikova M, Esteller M. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. Epigenetics 2011;6:692-702.
67. Dedeurwaerder S, Defrance M, Calonne E, Denis H, Sotiriou C, Fuks F. Evaluation of the Infinium Methylation 450K technology. Epigenomics 2011;3:771-84.
68. Roessler J, Ammerpohl O, Gutwein J, Hasemeier B, Anwar SL, Kreipe H, Lehmann U. Quantitative cross-validation and content analysis of the 450k DNA methylation array from Illumina, Inc. BMC Res Notes 2012;5:210.
69. Stirzaker C, Taberlay PC, Statham AL, Clark SJ. Mining cancer methylomes: prospects and challenges. Trends Genet 2014;30:75-84.
70. Bird A. DNA methylation patterns and epigenetic memory. Genes Dev 2002;16:6-21.
71. Ehrlich M, Gama-Sosa MA, Huang LH, Midgett RM, Kuo KC, McCune RA, Gehrke C. Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. Nucleic Acids Res 1982;10:2709-21.
72. Kitsios GD, Zintzaras E. Genome-wide association studies: hypothesis-"free" or "engaged"? Transl Res 2009;154:161-4.
73. Touleimat N, Tost J. Complete pipeline for Infinium((R)) Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. Epigenomics 2012;4:325-41.
74. Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M, et al. DNA methylation profiling of human chromosomes 6, 20 and 22. Nat Genet 2006;38:1378-85.
75. Lienert F, Wirbelauer C, Som I, Dean A, Mohn F, Schubeler D. Identification of genetic elements that autonomously determine DNA methylation states. Nat Genet 2011;43:1091-7.
76. Wang D, Yan L, Hu Q, Sucheston LE, Higgins MJ, Ambrosone CB, et al. IMA: an R package for high-throughput analysis of Illumina's 450K Infinium methylation data. Bioinformatics 2012;28:729-30.
77. Jaffe AE, Murakami P, Lee H, Leek JT, Fallin MD, Feinberg AP, Irizarry RA. Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. Int J Epidemiol 2012;41:200-9.
78. Peters TJ, Buckley MJ, Statham AL, Pidsley R, Samaras K, R VL, et al. De novo identification of differentially methylated regions in the human genome. Epigenetics Chromatin 2015;8:6.
79. Kechris KJ, Biehs B, Kornberg TB. Generalizing moving averages for tiling arrays using combined p-value statistics. Stat Appl Genet Mol Biol 2010;9:Article29.
80. Pedersen BS, Schwartz DA, Yang IV, Kechris KJ. Comb-p: software for combining, analyzing, grouping and correcting spatially correlated P-values. Bioinformatics 2012;28:2986-8.

81. Butcher LM, Beck S. Probe Lasso: a novel method to rope in differentially methylated regions with 450K DNA methylation data. Methods 2015;72:21-8.
82. Sofer T, Schifano ED, Hoppin JA, Hou L, Baccarelli AA. A-clustering: a novel method for the detection of co-regulated methylation regions, and regions associated with exposure. Bioinformatics 2013;29:2884-91.
83. Kolde R, Martens K, Lokk K, Laur S, Vilo J. seqlm: an MDL based method for identifying differentially methylated regions in high density methylation array data. Bioinformatics 2016;32:2604-10.
84. Burgess S, Timpson NJ, Ebrahim S, Davey Smith G. Mendelian randomization: where are we now and where are we going? Int J Epidemiol 2015;44:379-88.
85. Zheng J, Baird D, Borges MC, Bowden J, Hemani G, Haycock P, et al. Recent Developments in Mendelian Randomization Studies. Curr Epidemiol Rep 2017;4:330-345.
86. Ikram MA, Brusselle GGO, Murad SD, van Duijn CM, Franco OH, Goedegebure A, et al. The Rotterdam Study: 2018 update on objectives, design and main results. Eur J Epidemiol 2017;32:807-850.
87. Psaty BM, O'Donnell CJ, Gudnason V, Lunetta KL, Folsom AR, Rotter JI, et al. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from 5 cohorts. Circ Cardiovasc Genet 2009;2:73-80.

CHAPTER 2

Homocysteine and DNA methylation: a review of animal and human literature

Pooja R. Mandaviya, Lisette Stolk, Sandra G. Heil

Mol Genet Metab. 2014 Dec;113(4):243-52

ABSTRACT

Homocysteine (Hcy) is a sulfur-containing non-protein forming amino acid, which is synthesized from methionine as an important intermediate in the one-carbon pathway. High concentrations of Hcy in a condition called hyperhomocysteinemia (HHcy) are an independent risk factor for several disorders including cardiovascular diseases and osteoporotic fractures. Since Hcy is produced as a byproduct of the methyltransferase reaction, alteration in DNA methylation is studied as one of the underlying mechanisms of HHcy-associated disorders. In animal models, elevated Hcy concentrations are induced either by diet (high methionine, low B-vitamins, or both), gene knockouts (*Mthfr*, *Cbs*, *Mtrr* or *Mtr*) or combination of both to investigate their effects on DNA methylation or its markers. In humans, most of the literature involves case–control studies concerning patients. The focus of this review is to study existing literature on HHcy and its role in relation to DNA methylation. Apart from this, a few studies investigated the effect of Hcy-lowering trials on restoring DNA methylation patterns, by giving a folic acid or B-vitamin supplemented diet. These studies which were conducted in animal models as well as humans were included in this review.

HIGHLIGHTS

- Hyperhomocysteinemia in animals is associated with high SAH and low SAM/SAH ratio, but changes in SAM were not consistent.
- SAM:SAH ratio is not a good proxy for DNA methylation levels in hyperhomocysteinemic animal models.
- Both diet- and genetically induced hyperhomocysteinemic animal models have altered methylation indicating homocysteine as a keyplayer.
- Global DNA methylation was not consistently altered in humans with hyperhomocysteinemia.
- Homocysteine-lowering trials did not result in a clear improvement of DNA methylation patterns in most studies.

1. INTRODUCTION

Homocysteine (Hcy) is a sulfur-containing non-protein forming amino acid, which occurs naturally in the blood plasma. It is biosynthesized as an intermediate in the one-carbon pathway from methionine, via two main cofactors: S-adenosylmethionine (SAM) and S-adenosylhomocysteine (SAH). SAM acts as an important cosubstrate and is used by the DNA methyltransferase enzymes in transferring methyl groups to the DNA. The product of this reaction, SAH, is then synthesized to Hcy in a reversible manner. The concentrations of Hcy are maintained by two routes; namely: the remethylation pathway, where Hcy is converted back to methionine, and the transsulfuration pathway, where Hcy is converted to cystathionine to form cysteine [1]. A basic illustration of the Hcy pathway is given in Fig. 1.

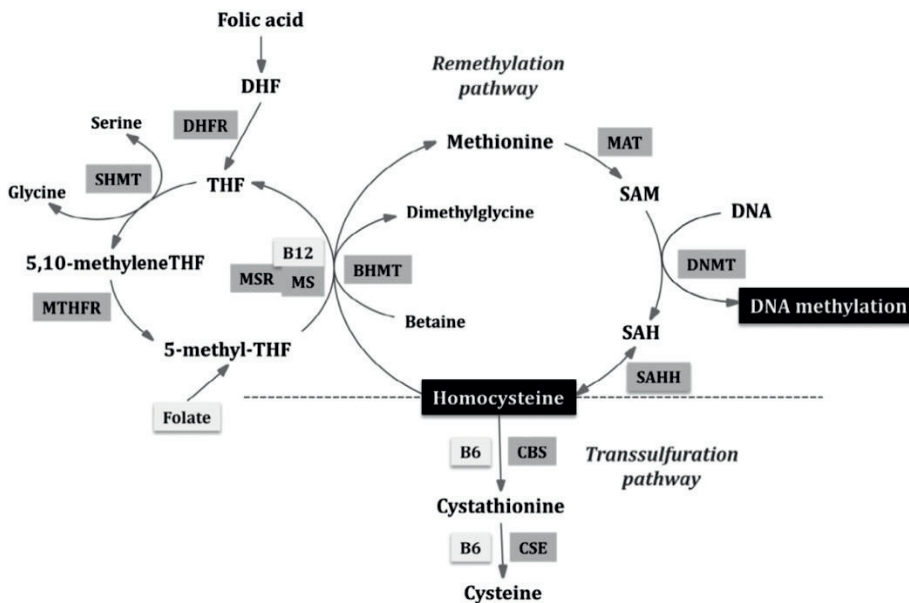


Figure 1. A simplified figure of homocysteine pathway as present in the liver, which includes the main enzymes and dietary factors described in this review. DHFR: Dihydrofolate reductase; SHMT: Serine hydroxymethyltransferase; MTHFR: 5,10-Methylenetetrahydrofolate reductase; MSR: Methionine synthase reductase; MS: Methionine synthase; BHMT: Betaine-homocysteine methyltransferase; MAT: Methionine adenosyltransferase; DNMT: DNA methyltransferase; SAHH: S-adenyl-l-homocysteine hydrolase; CBS: Cystathionine β-synthase; CSE: Cystathionine γ-lyase. Light gray boxes: B-vitamins; Dark gray boxes: enzymes; Black boxes: Homocysteine and DNA methylation; the dashed line divides the remethylation from the transsulfuration pathway.

Almost 20 years ago, high concentrations of Hcy in a condition called hyperhomocysteinemia (HHcy), were shown to be an independent risk factor for several disorders including cardiovascular diseases [2] and osteoporotic fractures [3]. Since the reaction from SAH to Hcy is reversible, high concentrations of Hcy increase the concentrations of SAH, which acts as a competitive inhibitor of the methyltransferase reaction [4]. Elevated SAH leads to lower SAM:SAH ratio, which could result in less donation of methyl groups to the DNA by SAM. Elevated SAH is shown to be associated with global DNA hypomethylation, but this phenomenon is tissue-specific and the mechanisms are unknown [5]. Since Hcy is produced as a byproduct of the methyltransferase reaction, alterations in DNA methylation levels are studied as one of the underlying mechanisms of HHcy-associated disorders.

The primary causes of HHcy are dietary and/or deficiencies of the key enzymes of the Hcy metabolism pathway [Fig. 1]. These dietary factors and enzymes that play a role in the remethylation and transsulfuration pathway, balance the concentrations of Hcy by converting it to methionine or cystathionine. The essential dietary factors include methionine and B vitamins (folate, vitamin B6 and vitamin B12). Low vitamin B12 and/or folate are associated with high Hcy [6] and associated risks like cardiovascular disease [7], pregnancy complications [8] and neural tube defects [9]. Besides methionine and B-vitamins, dietary choline and betaine are also important co-factors of the one-carbon pathway. Low dietary choline or with its association with dietary betaine is associated with elevated Hcy in both mice and humans [10–12]. Dietary choline undergoes oxidation to produce betaine and helps in the synthesis of SAM, thus being an indirect methyl donor contributing to the SAM:SAH ratio. Due to limited literature on choline deficient diets, we have focused this review only on high methionine (HM) and B-vitamin deficient diets.

Polymorphisms in genes encoding for enzymes like methylenetetrahydrofolate reductase (MTHFR), cystathionine β-synthase (CBS), and methionine synthase (MS) are important determinants of Hcy concentrations [13]. MTHFR converts 5,10-methylenetetrahydrofolate to 5-methyltetrahydrofolate (5-MTHF). 5-MTHF acts as a cosubstrate in the remethylation pathway by converting Hcy to methionine. The enzyme MS encoded by the MTR gene, is reductively activated by methionine synthase reductase (MSR) enzyme, which is encoded by MTRR. MS catalyzes the remethylation of methionine from Hcy. The enzyme CBS plays a role in the transsulfuration pathway by catalyzing the conversion of Hcy to cystathionine. Depending on the severity of one or more dietary and/or enzyme deficiencies, HHcy may occur at different levels. Hcy concentrations < 15 μmol/L are referred to as mild, between 15 and 30 μmol/L are referred to as moderate, between 30 and 100 μmol/L as intermediate, and >100 μmol/L as severe HHcy [14].

The focus of this review is to study existing literature on HHcy and its role in relation to DNA methylation. We included studies with effects of diet and/or genotype

on plasma or serum Hcy and its consequent role in the alteration of global, gene-specific or genome-wide DNA methylation. Scientific papers based on both animal and human experiments were reviewed in PubMed. A literature search with varied terms of Hcy and DNA methylation was done, in order to filter out relevant articles which were published from 2001 until 2014. In the animal literature, we focused on mice and rat studies. We included all papers which measured plasma or serum Hcy and either or all of the methylation markers like SAM and SAH or global or gene-specific DNA methylation. We divided the animal literature in three subtopics, namely 1) diet-induced HHcy, 2) genetically-induced HHcy, and 3) genetically- and diet-induced HHcy. A hyperhomocysteinemic (HH) diet mainly involves either HM, low folate (LF) or vitamin B12 or their combination. In addition, low concentrations of choline, riboflavin and pyridoxine might be present. From the human literature, studies which measured plasma or serum Hcy and DNA methylation levels were included, and only if the main aim of the study was to test the association between the two. We focused at vascular diseases, cancer, renal disease and brain disorders.

Methylation can be measured at 3 levels: global, genome-wide and gene-specific. We used the term “global” in this review for studies that report DNA methylation as the total 5-methyl cytosine content, using techniques like cytosine extension assay and LC–MS/MS. Some groups measured methylation of LINE-1, B1 and Alu repetitive elements, which are suggested to be surrogate markers of global methylation levels. We used the term “genome-wide” for methylation measures from DNA-methylation arrays. “Gene-specific” methylation measures methylation levels of cytosines located at specific genes or within their promoters.

2. ANIMAL STUDIES

2.1. Diet-induced HHcy

Diet with an excess of methionine and/or deficient in one or more of the B vitamins like folate, vitamin B6 or vitamin B12, play a role in elevating the concentrations of Hcy. The resulting deregulation of global or gene-specific DNA methylation and the methylation markers SAM and SAH is demonstrated with experiments involving different rat and mouse models [Table 1]. The apolipoprotein E (ApoE)- deficient mouse, a model that develops spontaneous atherosclerosis, have often been used in experiments to study the dietary impact of Hcy-induced DNA methylation alteration in vascular pathologies. A few other experiments are done in mice and rats without gene knock- outs to investigate similar effects.

2.1.1. High methionine diet

Among the many diseases linked to HHcy, vascular pathologies are mostly explored in the literature of Hcy and DNA methylation. In 2001, Dayal and associates showed in both Cbs (+/+) mice and Cbs (+/−) mice models from C57BL/6J background that, even in

the absence of folate deficiency, HHcy and endothelial dysfunction occur in mice fed with a HM diet for 15 weeks. SAM was lower in brain at 7 weeks and SAH was higher in liver at 15 weeks. Both in the brain and liver, SAM:SAH ratio was markedly reduced at 15 weeks, which had a strong correlation with high plasma Hcy [15].

In 2007, Jiang et al. investigated the effects of HHcy on DNA methylation of the B1 repetitive elements, in relation to atherosclerosis. They gave a HM diet with different concentrations of 1%, 1.5% and 2% to healthy Sprague–Dawley rats for 4 weeks to achieve HHcy. B1 elements, as quantified by real-time PCR, were hypomethylated to 26.2%, 20.1% and 22.2%, respectively for the three diet concentrations. Aortic SAM decreased and SAH increased, leading to a 3- to 4-fold decrease in SAM:SAH ratio [16]. In order to further examine these effects of HHcy on the pathogenesis of atherosclerosis more thoroughly, the same group also studied an ApoE-deficient mouse model. For 15 weeks, the mice were given a HM diet to induce HHcy. Methylation of both global DNA and B1 elements was measured in the aorta, by both methylation-dependent restriction analysis and nested methylation- specific PCR (MSP). The HM group showed a lower methylation of global DNA and B1 elements, compared to controls. Aortic SAM and SAH were 1.35- and 1.86-fold higher, respectively. SAM:SAH ratio showed an opposite trend of being 1.46-fold higher unlike previous study, and the atherosclerotic lesions were also larger [17]. In a similar study, ApoE-deficient mice when given a HM or HM-B-vitamin deficient diet showed an increase in plasma Hcy in the HM diet group, with a more remarkable increase in HM-B-vitamin deficient group. Plasma SAH was higher in both groups, compared to the ApoE-deficient mice with control diet. The atherosclerotic lesion areas were 53% and 95% larger in the HM and HM-B-vitamin deficient groups, respectively. Aortic global methylation identified by detection antibodies and quantified by an ELISA, was lower in both groups [18]. Another similar study examined HHcy effects on cardiac injury by gene-specific DNA methylation alteration in ApoE-deficient mice induced by a HM diet. Lower cardiac SAM and 5.2% lower methylation of the key cardiac apoptotic gene, Trp53, was observed compared to the ApoE-deficient mice with regular diet, as measured by nested MSP. It remains unclear what the effect of these small changes in methylation is on cardiovascular phenotypes of these mice [19].

In addition to vascular pathologies, methionine-induced HHcy is explored in relation to other disorders like bone loss and neurodegeneration. A HM diet given for 20 weeks to female Wistar rats, induced HHcy and accumulated Hcy in bone tissue by 13 times leading to a reduced bone strength and cancellous bone loss. SAM and SAH concentrations were increased and the SAM:SAH ratio was decreased in the plasma and bone [20]. The same authors in a later study gave a HM diet to female Wistar rats for 5 months to study the effects of HHcy in relation to neurodegeneration. As a result, SAH was higher and SAM:SAH ratio was lower in plasma, brainstem and frontal cortex of rats on HM diet, compared to the rats on control diet. SAM concentrations were

higher in the plasma and brainstem [21].

2.1.2. B vitamin deficient (folate, B6, B12) diets

Similar to HM diet, a B-vitamin deficient diet to induce HHcy in animal models has been studied too. Caudill et al. administered a diet low in methionine and deficient in folate and choline to C57BL/6J mice. After 24 weeks, Hcy was higher, compared to the control diet mice. SAM concentrations and SAM:SAH ratio were lowered in all the analyzed tissues; liver, kidney, brain and testes. SAH was higher in the liver, but not in other tissues. However, global DNA methylation measured by cytosine extension assay did not change in any of the tissues [22].

In the context of vascular pathologies, effects of HHcy induced by a B-vitamin deficient diet were not as clear as compared to the effects by the HM diet. In a study by Liu et al., B-vitamin deficient diet with or without HM was given to a group of ApoE-deficient mice. The aortic sinus plaque areas were shown to be larger as a result of HM diets, but not with B-vitamin deficient diet alone. Plasma Hcy was elevated, but there was no change in the plasma SAH or aortic global methylation levels [18]. Another study investigated the effects of folate and/or vitamin B12 deficiency on endothelium-dependent relaxation in rats. Serum Hcy was higher in the folate deficient (FD) and folate-vitamin B12 deficient diet groups, compared to controls. Liver SAM and SAM: SAH ratio were lower, and SAH showed no difference. However no association of folate, vitamin B12, Hcy, SAM, SAH or SAM:SAH ratio was found with vascular reactivity [23]. In yet another study, moderate HHcy resulted when ApoE-deficient mice were treated with either a FD or folate, vitamin B6 and B12 deficient diet for 16 weeks. The hepatic SAM:SAH ratio for both the diets were reduced up to 80 and 90% respectively. However, they showed no association with atherosclerosis and global DNA methylation of the vascular or liver tissue, even after an increased atherosclerotic lesion formation in the aortic arch [24,25].

In relation to pregnancy outcomes, maternal folate status affects the homeostasis of Hcy pathway in their offsprings. In a study by Blaise et al., female Wistar rats were fed a diet deficient in vitamin B12, B2, folate and choline from one month before pregnancy until weaning at day 21. Hcy concentrations were moderately elevated in mothers. The pups, who were fed on dams with such a deficient diet, also developed HHcy and had a decline in hepatic SAM:SAH ratio due to the decrease in SAM. Cbs, Mthfr and Ms activities in the liver were extremely lower in deficient pups in comparison to normal fed pups [26]. In another similar study of a FD diet given to female pregnant rats starting 2 weeks before mating until gestation at day 21, 4.5-fold increased maternal plasma Hcy concentrations were seen. However in this case, no change in the global DNA methylation levels was observed in the maternal or fetal rat livers [27]. Similarly, Mejos et al. randomly administered a FD or a folate supplemented (FS) diet to male and female rats for 4 weeks until mating. Hcy concentrations were

elevated in the postnatal rats of either or both of the deficient parents, in comparison to the postnatal rats of both FS parents. There was a marked reduction of hepatic folate and global DNA methylation levels as identified by detection antibodies and quantified using an ELISA [28]. In another study, pregnant Mthfr (+/+) mice were fed a FD diet to induce HHcy. Hepatic SAM concentrations were decreased and SAH concentrations were increased, which was accompanied by a reduction of SAM:SAH ratio. Placental SAH also increased with decrease in SAM:SAH ratio [29].

Two other studies measured the levels of DNA methylation and its metabolites in the brain of FD diet induced HH rats. The treatment periods for the studies were 30 days and 36 weeks. SAM, SAH and SAM: SAH ratio remained unaltered in both cases. In the first case, global DNA methylation measured by in vitro methyl acceptance capacity assay, was lowered in the deficient group compared to controls [30]. Interestingly, in the second case, at 18 and 36 weeks, the global DNA methylation levels quantified by high-performance liquid chromatography–(electrospray)-mass spectrometry, were higher in the FD diet group, as compared to controls, although the same mice group had also shown global hypomethylation in the liver, when quantified by the cytosine extension assay [31,32]. The trend in the alteration of DNA methylation differed between studies, or might suggest a different mechanism with prolonged treatment period. Furthermore in an additional study, such a similar HH model of Cbs (+/+) mice showed elevated SAH with lower SAM:SAH ratio in the brain [33].

Folate deficiency, which is also implicated in relation to colon cancer, was experimented in the colon of HH rats by two groups. In the first study, SAM concentrations of the colonic mucosa and methylation of Trp53 promoter region remained unchanged in the FD diet group. A 3–3.5 fold higher SAH and 64–71% lower SAM:SAH ratio was observed at 5 weeks, in comparison to controls. A 30% higher colonic global DNA methylation, measured by an in vitro methyl acceptance capacity assay was observed only at 3 weeks, which directly correlated with plasma Hcy [34]. The second study group conducted a FD treatment of 24 weeks. An ~11% lower hepatic SAM:SAH ratio was observed, in comparison to controls. But no change was found in colonic SAM and SAH concentrations or their ratio. Global DNA methylation in the liver and colon did not lower significantly, even after this longer treatment period of 24 weeks [35].

In general, different diets may have variable effects on Hcy, DNA methylation and its markers. Devlin and associates conducted a study, where they assigned one of the 3 diets for 7–15 weeks to random groups of both Mthfr (+/+) and Mthfr (+/−) mice; 1) HM 2) LF 3) HM-LF. Plasma Hcy showed no difference between control and the HM diet group. But higher Hcy was observed in the LF and HM-LF group, as compared to controls. Hepatic and brain SAM:SAH ratio were lower in all experimental diet groups, being the most in the LF and HM-LF groups. In the liver, the lower SAM:SAH ratio in LF and HM-LF groups was due to higher SAM, whereas, in the HM group, it was due to

higher SAH. No effect on global hepatic or brain DNA methylation was observed in any of the three cases [36].

2.2. Genetically-induced Hhcy

Genetic variants in the key enzymes of the one-carbon pathway like MTHFR, MTRR, MS or CBS are associated with disease risks. These enzymes help balance Hcy concentrations within the pathway. In order to study Hcy-induced change in DNA methylation or its markers by such enzyme deficiencies, a few study groups used gene knockout mice models [Table 2].

2.2.1. MTHFR deficiency

MTHFR is one of the common enzymes that could be deficient leading to alterations in Hcy metabolism. In relation to vascular pathology, Chen et al. generated two groups of Mthfr (+/−) and Mthfr (−/−) mice, which showed 1.6 and 10-fold higher plasma Hcy, respectively, compared to Mthfr (+/+) mice. SAM and SAH concentrations in the brain, ovaries, testes and liver were altered, and global DNA methylation, measured by cytosine extension assay, was lowered in the brain and ovaries. This was accompanied by an abnormal lipid deposition in the aorta for both knockout groups, and a 76.4% reduced survival of Mthfr (−/−) mice [37]. In 2004, Devlin et al. used the Mthfr (+/−) mice that were generated by Chen et al. They observed lower SAM: SAH ratio in the liver and brain tissue compared to Mthfr (+/+) mice. Disrupted Mthfr gene also sensitized the mice to diet-induced HHcy and endothelial dysfunction [36]. In yet another study, Mthfr (−/−) mice showed lower global DNA methylation which was quantified by thin layer chromatography in the cerebellum, as compared to the control mice. They also found lower methylation of CpG dinucleotide 4 within first exon 1_7 of the Nr3c1 gene encoding the glucocorticoid receptor in the hippocampus, which was measured by pyrosequencing. This was accompanied by more mRNA and protein of the Nr3c1 gene, behavioral anomalies, disrupted cognitive function, increased apoptosis and other neurobiological changes [38].

2.2.2. CBS deficiency

In 2001, Caudill and coworkers studied the effects of HHcy on Cbs (+/−) mice. As a result, SAM and SAM:SAH ratio were lower in all the analyzed tissues; liver, kidney, brain and testes, as compared to the Cbs (+/+) mice. SAH showed an increase in the liver and testes, but not in the kidney and brain. DNA methylation was measured by cytosine extension assay and did not show to be hypomethylated in any of the tissues, but the testes [22]. In the same year, Dayal et al. set up a similar Cbs (+/−) mouse model to induce HHcy. In the liver and brain, SAH were 2-fold higher and SAM:SAH ratios were lowered. In addition, an approximately 20% lower thrombomodulin activity was observed in the aortic arch of the Cbs (+/−) mice, which correlated with SAH [15]. Sontag et al. in their defined HH model also showed an ~27% lower methylation of

protein phosphatase 2, regulatory subunit A, alpha (Ppp2r1a) gene in the brain. The methylation was identified by detection antibodies and quantified by ELISA [33]. In three studies that followed, severe HHcy showed in Cbs (−/−) mice. In the first study, SAH concentrations were 8 times higher in the liver and kidney and 190 times higher in the brain, as compared to Cbs (+/+) mice. The SAM:SAH ratio was observed to be 80% in the liver, 87% in the kidney and 100% in the brain. Only in the liver tissue, SAM concentrations were higher and global DNA methylation measured by the in vitro methyl acceptance capacity assay was lower [39]. In the second study, 20-fold higher liver Hcy in 4- to 6-month old Cbs (−/−) mice were seen, as compared to Cbs (+/+). Hepatic SAM and SAH concentrations were also higher [40]. In the third study, Cbs (−/−) mice developed fatty livers at 2 weeks, which became less apparent at 8 weeks. The fatty livers at 2 weeks also showed a decline in global protein arginine methylation. The liver SAM and SAH concentrations increased at both weeks, showing a 2-fold decrease of the SAM:SAH ratio at 8 weeks [41].

2.2.3. MSR deficiency

The enzyme MSR is required for the normal functioning of the MS enzyme, which helps in Hcy remethylation. In 2005, Dayal et al. com- pared Mtr (+/−) mice with Mtr (+/+) mice, and demonstrated that higher plasma Hcy was observed in the Mtr (+/−) mice, compared to the Mtr (+/+) mice. However, no effect was seen on methylation markers of the liver or brain tissue [42]. In 2007, Elmore et al. generated HH mice with either a heterozygous or a homozygous knockout for the Mtrr gene, encoding MSR. No difference in SAM and SAH was observed in the kidney and brain. Heart tissue showed higher SAH and lower SAM:SAH ratio, as compared to the Mtrr (+/+) mice. But unexpected opposite trends were observed in the liver, which showed higher SAM and lower SAH with higher SAM:SAH ratio [43] study demonstrated that maternal Mtrr (−/−) mice influences the placental methylation levels of their Mtrr (+/+) grandprogeny, both globally and in about 45–70% of the studied imprinted genes. In this case, global DNA methylation was identified using detection antibodies and methylation of the imprinted genes was quantified by bisulfite pyrosequencing [44].

2.3. Genetically- and diet-induced Hhcy

Dietary deficiencies affect Hcy concentrations and consequently DNA methylation. However if dietary deficiencies are accompanied by enzyme deficiencies, then a more sensitive effect on Hcy and DNA methylation can be expected [Table 3]. A few studies in rat and mouse models have shown a combined effect of diet and gene knockout to investigate the same.

2.3.1. Diet and CBS deficiency

Caudill et al. studied the HHcy effects from a diet deficient in methionine, choline and folate, in a Cbs (+/−) mice. SAM and SAM:SAH ratio were lowered in all analyzed

tissues; liver, kidney, brain and testes, compared to Cbs (+/+) mice with regular diet. SAH concentrations were higher in the liver, brain and testes, but not in the kidney. Using the cytosine extension assay, global DNA was found to be hypomethylated only in the brain and testes [22].

In relation to vascular pathologies, several studies administered a HM-LF diet to Cbs (+/−) mice to induce HHcy. Using such a mouse model, tissue-specific alterations of methylation in the maternal differentially methylated domain of the H19 gene were observed. This domain was hypomethylated in the liver, and hypermethylated in the brain and aorta, as compared to the Cbs (+/+) mice with control diet. SAH concentrations increased and SAM:SAH ratio decreased in the liver and brain. The expression of H19 transcripts in the aorta increased by 2.5-fold [45]. In a similar study, when giving a HM-LF diet to Cbs (+/−) mice, Glier et al. reported mildly higher Hcy. SAH concentrations were higher and SAM:SAH ratio was lower in the liver, but not in the brain, as compared to Cbs (+/+) mice with control diet. Lower SAM:SAH ratio in the liver was accompanied by lower methylation in the maternal H19 DMD allele which was measured by bisulfite sequencing. Interestingly, higher methylation in this domain was observed in the brain, despite no effect in brain SAM:SAH ratio. This also suggests tissue-specific differences [46]. Previously, the authors observed alteration of the SAM and SAH in the liver and brain at 7 and 15 weeks of diet treatment, with a lower SAM:SAH ratio at 15 weeks [15]. In another recent study, 12 weeks of HM-LF diet given to a Cbs (+/−) mice showed hypermethylation of the promoter transcripts of the Nr3c1 gene, measured using bisulfite pyrosequencing. This epigenetic silencing of Nr3c1 gene in the aorta was associated with HHcy [47].

With a combination of Cbs knockout and HM-LF diet in mice, studies have also been conducted in relation to HHcy associated liver pathologies. In the first study, a fatty acid desaturase enzyme of the liver, which is encoded by Fads2 gene and involved in fatty acid synthesis, was epigenetically silenced with hypermethylation in its promoter in Cbs (+/−) mice with HM-LF diet, compared to Cbs (+/+) mice with control diet. This was accompanied by lower SAM:SAH ratio, which associated with phospholipid and fatty acid metabolism in the liver [48]. In a similar study, the liver acyl CoA:cholesterol acyl transferase 2 (Acat2) enzyme responsible for the production of cholesteryl esters, had lower expression levels in Cbs (+/−) mice with HM-LF diet, compared to Cbs (+/+) mice with regular diet. Soat2 which encodes Acat2, had an altered methylation which was quantified by bisulfite pyrosequencing at the CpG rich region of its 5′ portion, compared to Cbs (+/+) mice with control diet. This was also accompanied by higher methylation of the B1 repetitive elements that were quantified by the same technique [49].

3. HUMAN STUDIES

Besides in animals, effects of Hcy on DNA methylation are also investigated in humans. Most of these concerned patients and a few healthy volunteers [Table 4]. In a study, 71 vegetarians, 58% of whom were vitamin B12 deficient, were evaluated for their methylation markers and their relationship with peripheral blood DNA methylation. Serum Hcy was higher than 12 µmol/L in 45% of the vegetarians, which was higher than in omnivores. It showed no association with SAM, SAH or SAM:SAH ratio. Methylation levels of 3 CpG sites in Alu-repetitive elements correlated with SAH and SAM:SAH ratio, but not with Hcy or SAM. In addition, the methylation status of the SHC1 promoter was investigated, however this was not correlated with Hcy, SAM, SAH or SAM:SAH ratio [50].

In another study, healthy adult women with elevated Hcy had a 2.6-fold lower lymphocyte global DNA methylation, which was measured by cytosine extension assay. Plasma SAH concentrations were 2-fold elevated, which decreased SAM:SAH ratio by one-half. This SAM:SAH ratio showed a direct relation with intracellular lymphocyte SAM:SAH ratio [51,52]. In a study by Perng et al. plasma Hcy positively correlated to LINE-1 methylation, which was quantified by pyrosequencing in healthy middle-aged adults. However, Alu methylation did not significantly correlate with plasma Hcy [53].

A study by Farkas and group divided Hcy concentrations of 56 blood leukocyte samples into low (5–10 µmol/L) and high (20–113 µmol/L) Hcy and considered them as well-nourished and poorly nourished subjects, respectively. Site-specific methylation differences quantified by pyrosequencing were observed in 17 CpG sites of the RFC1 gene between the groups. [54].

In relation to pregnancy, LINE-1 methylation in blood and folate- related parameters from 24 pregnant women, all of whom were under folate diet, were determined using pyrosequencing. Cord plasma Hcy showed inverse correlation with LINE-1 methylation in cord blood [55]. This study was later extended by using DNA methylation arrays in 12 pregnant women in order to assess correlation of genome-wide methylation patterns with plasma Hcy and LINE-1 methylation. Univariate regression analysis identified 298 autosomal and 8× chromosome CpG sites associated with Hcy. In addition, methylation of 146 autosomal CpGs mostly located within CpG islands, was associated with LINE-1 methylation [56].

A few more studies were performed in individuals having specific diseases, which have known to be associated with high Hcy. We focused this review at vascular disease, cancer, renal diseases and brain disorders.

3.1. Vascular disease

Two case-control studies were performed to evaluate Hcy effects on global DNA methylation in vascular diseases, measured using cytosine extension assays. The first study measured global DNA methylation in leukocytes, which was lower in

atherosclerotic patients (N = 17), as compared to controls (N = 15) and had an inverse correlation with plasma Hcy. In addition, plasma SAH concentrations were higher and SAM:SAH ratio was lower, which also correlated with Hcy [57]. The second study measured global DNA methylation in peripheral blood lymphocytes. In this case, Hcy and global DNA methylation were higher in 137 coronary artery disease cases compared to 150 controls, with a positive association between the two [58]. In another study, LINE-1 methylation was measured at baseline in 712 elderly men either at low risk or with high risk for ischemic heart disease or stroke. Plasma Hcy was higher and LINE-1 methylation was lower in the case of individuals at risk. However, the two were not correlated [59].

DNA methylation was further studied in a few candidate gene studies. Two case–control studies with atherosclerotic patients were performed. In the first study (54 cases, 28 controls), elevated Hcy concentrations positively correlated with DNA methylation of CpG islands at the estrogen receptor alpha (ESR1) promoter, which were quantified using nested MSP [60]. In the second study (197 cases, 165 controls), elevated Hcy was correlated with MSP measured DNA hypomethylation of human telomerase reverse transcriptase (hTERT) and its downregulation, which was accompanied by shortened leukocyte telomere length. This was also confirmed in methionine fed HH mouse model [61]. High Hcy was also inversely correlated with site-specific methylation of the SHC1 promoter which was quantified using Sequenom EpiTyper in peripheral blood leukocytes of coronary artery disease patients. [62] . In a group of 192 patients that underwent coronary artery bypass grafting, there was no correlation between Hcy and DDAH2 methylation of the internal mammary artery, but with NOS2 methylation of the internal mammary artery and punch aortic fragments [63].

In a study by Heil and group, heparin blood samples were used to isolate genomic DNA from 9 CBS deficient patients treated with Hcy lowering therapy and 8 healthy controls. Hcy, SAM and SAH were higher in patients than in controls. However, no difference in SAM:SAH ratio was found. In addition, global DNA methylation and gene- specific methylation of H19 DMR were measured, which showed no difference compared to controls [64].

3.2. Cancer

In 2003, Pufulete et al. observed 26% lower colonic global DNA methylation in patients with colorectal adenoma (N = 35) and 30% lower methylation in 28 patients with colorectal cancer, as compared to 76 controls. DNA methylation was quantified using in vitro methyl acceptance capacity assay. Plasma Hcy, even though not significantly higher in patients, showed an association with colonic DNA methylation [65]. The same group also generated methylation indexes of the ESR1 and mutL homologue 1 (MLH1) genes by PCR and pyrosequencing in the same subjects, which were higher in patients

compared to controls. The methylation indexes of the ESR1 promoter correlated positively with Hcy, while those at MLH1 did not show any correlation [66]. Similarly, a colorectal cancer case–control study investigated the relation between variables of one-carbon metabolism and promoter methylation of an eight-gene panel (CDKN2A, MLH1, CACNA1G, NEUROG1, RUNX3, SOCS1, IGF2, and CRABP1) of archival colorectal tumor tissue in colorectal cancer cases (N = 190). The methylation of these genes did not associate with Hcy [67].

In summary, these studies demonstrate that MLH1 gene methylation was not correlated with Hcy in colorectal cancer patients. However, in patients with esophageal diseases (N = 124), promoter MLH1 hyper- methylation as quantified by restriction enzyme-based methylation analysis, was observed in 63.5% of cancer cases and in 53.8% of precancer cases, which showed association with elevated plasma Hcy [68].

Two further studies were conducted in breast cancer patients, where in promoter methylation of the targeted genes was quantified by MSP. In the first study, the RARB promoter and exon 1 regions of ESR1 in breast tumor tissue of 137 primary breast cancer cases were hypermethylated and showed elevated Hcy [69]. In the second study (179 cases), elevated Hcy correlated with BRCA1 promoter hypermethylation in the breast cancer tissue and corresponding peripheral blood of all breast cancer types. In addition, RASSF1 promoter hypermethylation was studied in breast cancer tissue of estrogen receptor (ER)-negative and progesterone receptor (PR)-negative breast cancer cases and was also shown to correlate with Hcy [70].

3.3. Renal disease

Ingrosso et al. investigated the effects of HHcy on global DNA methylation in 32 patients with uremia who had elevated Hcy, and compared them with 11 healthy controls. DNA methylation of peripheral blood mononuclear cells (PBMCs) measured by cytosine extension assay and southern blotting was lower in patients, which associated with HHcy. This change in global DNA methylation was associated with loss of imprinting of the H19 DMR [71]. In two other studies, global DNA methylation was measured in leukocytes of 78 stages 2–4 chronic kidney disease or 20 chronic hemodialysis patients. However, in these studies, global DNA methylation was not associated with plasma or serum Hcy [72,73].

3.4. Brain disorders

Patients with Alzheimer's disease showed a negative association of plasma SAM:SAH ratio with plasma Hcy. In addition, SAM and SAH concentrations were higher in patients [74]. Case–control studies were per- formed with euthymic bipolar or schizophrenia patients by Bromberg et al. Hcy was elevated in schizophrenia patients, but not in euthymic bipolar patients compared to controls. Global methylation levels in peripheral blood leukocytes did not differ between either of the cases and controls,

and was not associated with Hcy [75,76]. A similar study performed in schizophrenia patients (N = 42), examined the effect of plasma Hcy on genome-wide DNA methylation of peripheral leukocytes measured by arrays. Patients had moderate concentrations of Hcy, as compared to controls. Out of the analyzed CpGs, 1338 of them, annotated at different gene regions, showed to be differentially methylated with elevated Hcy. 56.6% of these sites were located in the CpG islands and their flanking regions. Two of them included the methylated sites from the SLC18A2 and GNAL genes, which had previously shown to associate with schizophrenia. However no correction for multiple testing was used [77].

3.5. MTHFR 677C>T

Multiple studies looked at differences in Hcy and DNA methylation comparing MTHFR 677CNT (rs1801133) CC and TT genotypes. Two studies investigated levels of global DNA methylation, one in PBMCs using LC Ion Trap LC/MS and the other in leukocytes using cytosine extension assay. In both studies, the MTHFR TT genotype had lower levels of methylation, in comparison to subjects carrying the CC genotype, which positively correlated with folate status and negatively correlated with plasma Hcy [78,79]. Two other studies were conducted in young women who participated in a regime of folate depletion and repletion diet for 2 consecutive periods of 7 weeks. Global DNA methylation in leukocytes was measured at baseline, weeks 7 and 14. In the first study, methylation levels did not decrease significantly after 7 weeks of folate depletion. But at 14 weeks, the methylation increased only in women with the TT genotype. Folate concentrations were reduced and plasma Hcy were elevated in women with TT genotype, both post-depletion and post-repletion. Hcy directly correlated with DNA methylation only during depletion [80]. In the second study, there was no difference in methylation levels post-depletion, but they were lower after post-repletion. A delay in methylation response to folate in- take was suggested here, and no correlation was observed with Hcy [81].

Three other studies assessed methylation in controls of colorectal cancer subjects according to their genotypes. However, no difference in global DNA methylation of lymphocytes, colon or colonic mucosa was observed when stratifying between MTHFR CC and TT genotypes. Also, no association was found between Hcy and methylation [82,83], except in colonic mucosa, where plasma Hcy showed a 110% increase in individuals of TT genotypes [84].

4. EFFECT OF HCY-LOWERING THERAPY IN RELATION TO METHYLATION

4.1. Animal studies

Few studies investigated the effect of Hcy-lowering therapy on DNA methylation by giving a B-vitamin supplemented diet to either rat or mouse models [Table 5]. Three studies investigated the effect on liver DNA methylation. A study using male Wistar rats

given a 40 mg/kg high folate diet for 4 weeks, showed lower plasma Hcy, compared to controls. But no change was observed in hepatic SAM, SAH, SAM:SAH ratio or global DNA methylation [85]. Another study fed Sprague–Dawley rats a vitamin B12 deficient diet for 6 weeks to induce high Hcy. In addition, these rats were simultaneously given a high folate diet of 100 mg/kg. Similar to the previous study, Hcy concentrations were normalized by this treatment, but there was no change in hepatic and brain SAM, SAH, SAM:SAH ratio or hepatic global DNA methylation [86]. Later, when Sie et al. investigated the effect of high maternal and post-weaning folate diet on the offspring of female Sprague–Dawley rats, contrasting results were observed. At weaning, hepatic global DNA methylation which was quantified by in vitro methyl acceptance capacity, was decreased by 25% due to maternal folate diet of 5 mg/kg. Hcy and site-specific methylation of Pparg, Esr1, Trp53 and Apc genes measured by methylation-sensitive restriction enzyme, were also decreased. At 14 weeks of post-weaning, Hcy decreased due to both maternal and post-weaning folate diet, and methylation levels of Pparg, Trp53 and P16 genes were increased. A decrease in global DNA methylation and increase in methylation of Esr1 and Apc genes were also observed, but only due to post-weaning supplementation [87].

In the case of colon tissue, male Sprague–Dawley rats on a 8 mg/kg high folate diet for 5 weeks, showed no modulation of SAM, SAH, SAM:SAH ratio or DNA methylation at any measured time-point, as compared to controls [34]. Yet again, contrasting results were seen when Sie et al. in 2011 investigated maternal and post-weaning folate diet effects on colonic global DNA methylation quantified by liquid chromatography–electrospray ionization mass spectrometry. Hcy at weaning decreased due to maternal high folate diet of 5 mg/kg, and colonic global DNA methylation increased by 3%. At 14 weeks of post-weaning, Hcy similarly decreased due to both maternal and post-weaning supplementation, but colonic global DNA methylation decreased only due to post-weaning supplementation [88].

For brain tissue, 30 days of 8 or 40 mg/kg high folate diet did not modulate brain methylation markers [30]. Another recent study used a Cbs (+/−) mouse model on a HM diet for 4 weeks to induce HHcy, which increased brain DNA methylation. To investigate the effects of high folate diet, they supplemented 0.0057 µg/g bodyweight per day dose of folic acid in drinking water to these mice. As a result, Hcy concentrations were lowered in the HHcy mice, and brain methylation was lower, although not significant. Global DNA methylation was detected using antibodies and quantified by ELISA [89].

4.2. Human studies

In humans, Hcy-lowering therapy studies did not lead to promising results for restoring DNA methylation patterns [Table 6]. In case of colorectal adenoma patients, Pufulete and coworkers performed a placebo-control trial with a short-term folate diet given

400 μg/day to patients with colorectal adenoma for 10 weeks. As a result, 12% lower plasma Hcy was observed. Successively, global DNA methylation measured by in vitro methyl acceptance capacity in leukocytes and colonic mucosa were higher by 31% and 25%, respectively, suggesting that DNA methylation can be reversed by this therapy [90]. The same authors also investigated whether 400 μg/day folate diet re- stores methylation in CpGs of the ESR1 and MLH1 genes of colonic mucosa, since they were previously shown to be hypermethylated in case of colorectal tumors. After supplementation, Hcy decreased in folate diet group, but no effect was found on ESR1 and MLH1 methylation [91]. Another placebo-control study was conducted with a 6 months daily supplementation of 5 mg/day folate diet and 1.25 mg/day vitamin-B12 to patients with previous colorectal adenomas. Again, Hcy concentrations were lower. However no methylation change was determined in the promoter of 6 tumor suppressor and DNA repair genes in rectal biopsies [92].

Ingrosso et al. gave a folate treatment of 15 mg/day oral methyltetrahydrofolate to patients with HHcy and uremia for 2 months, as a result, plasma Hcy was decreased. Global DNA of PBMCs which was hypomethylated before treatment, restored methylation levels showing an increase. This was measured by cytosine extension assay and southern blotting. In addition, folate was able to restore loss of imprinting of the H19 gene [71]. When giving 5 mg/day folate diet to patients with intermediate HHcy for 8 weeks, Hcy concentrations were restored with an increase in plasma SAM and SAM:SAH ratio. However, no change in plasma SAH or global methylation levels of PBMCs was observed [93].

End stage renal failure patients have increased global DNA methylation in peripheral lymphocytes. In order to investigate whether Hcy-lowering therapy could reduce lymphocytic methylation, long- term stable dialysis patients were given a short term folate diet (15 mg thrice weekly) with or without an additional supplementation of vitamin B12 (1000 μg once a week). At 12 weeks, plasma Hcy was lower in folate diet group, compared to controls. However, lymphocyte global DNA methylation measured at 20 weeks, did not change [94]. Also a 1 year Hcy-lowering treatment with B-vitamins (folate diet: 5 mg/day, pyridoxine: 100 mg/day and vitamin B12: 1 mg/day) in stage 2–4 chronic kidney disease patients, showed no change in global DNA methylation in leukocytes [72].

Three other long-term studies were conducted for either 1, 2 or 3 years. In the first study, the effect on whole blood LINE-1 methylation was studied by supplementing adult subjects with B (folate diet: 500 μg/day, vitamin B12: 500 μg and vitamin B6: 50 mg) and D vitamins (1200 IU) for 1 year. As a result, no change in LINE-1 methylation levels was observed, but the plasma Hcy decreased [95]. In the second study, a Hcy-lowering therapy of 1 mg/day folate and B vitamins (vitamin B12: 500 μg/day and vitamin B6: 10 mg) given to an older healthy group with elevated Hcy of N13 μmol/L showed a reduction in plasma Hcy after 2 years, while no difference was

found in the plasma SAM, SAH and the SAM:SAH ratio [96]. On the third study, a placebo-control trial of participants with moderately elevated Hcy between 13 µmol/L and 26 µmol/L was investigated with a daily supplementation of 0.8 mg/ day folate diet for 3 years. After supplementation, plasma Hcy decreased by 21.6%. However, no difference in global DNA methylation in peripheral blood leukocytes was observed between the treatment groups [97].

In addition to therapies, a recent study investigated the impact of folic acid fortification of the US food supply on leukocyte global DNA methylation in 408 postmenopausal women. After nearly 2 years of fortification, they observed a reduction in plasma Hcy and increase in RBC folate. There was no change in leukocyte global DNA methylation. However, when these women were grouped into highest and lowest tertiles of RBC folate distribution, they observed that the group with the highest tertile had elevated mean DNA methylation in the pre-fortification period and lower mean DNA methylation in the post-fortification period, in comparison to the lowest tertile group. Global DNA methylation was negatively correlated with plasma Hcy in the pre-fortification period and positively in the post-fortification period. Thus RBC folate and plasma Hcy were proved to be important predictors of leukocyte global DNA methylation [98].

5. DISCUSSION AND CONCLUSION

We reviewed the current literature on Hcy and DNA methylation. Although in most studies, a clear association between Hcy and DNA hypomethylation was observed, results were not always consistent. In animals, the association between Hcy and DNA hypomethylation was clear whereas in humans, several contradictory findings were published. This might be caused by the pathophysiological effects of the underlying disease or might be explained by genetic differences between humans.

HH animal models in general have higher SAH, lower SAM:SAH ratio and lower global DNA methylation compared to normohomocysteinemic animals. However, a few studies showed higher SAM [15,20,21]. In addition, Cbs (−/−) mice with severe HHcy also showed higher SAM concentrations in the liver with lower SAM:SAH ratio [39,41]. Despite higher SAM concentrations, lower SAM:SAH ratio might be the result of severely elevated SAH, which is due to severely higher concentrations of Hcy in Cbs (−/−) mice. Hence, severely elevated SAH concentrations act as a strong inhibitor of methyltransferase reaction, despite of the higher SAM. On the contrary, two other studies in Mtrr (+/−) and Mtrr (−/−) mice found higher SAM, lower SAH and higher SAM:SAH ratio in the liver [17,43,44]. This suggests that different genetic backgrounds influence the outcomes differently.

SAM:SAH ratio, also called the methylation potential, did not seem to be a proxy for DNA methylation levels in HH animal models. Some studies showed a similar trend between the two, but in other cases, a clear opposite trend was observed. Jiang Y

et al. observed higher aortic SAM:SAH ratio with lower global DNA methylation [17], while Devlin AM et al. observed a lower brain and hepatic SAM:SAH ratio with higher global DNA methylation [45,48].

Hcy concentrations did not seem to be dependent on the type of treatment i.e. by diet, gene knockout or both. However, three studies which investigated both independent and combined effects of diet and genotype, observed higher Hcy concentrations in HH genetic model, a similar or a higher effect in HH diet model, and a markedly higher effect in the combined model, as compared to the control model [15,22,36]. Despite these difference in Hcy, SAM:SAH ratios did not always show a clear difference between the three models. Hence, the association between Hcy and methylation markers seems independent of the underlying cause; i.e. effects were consistently shown in diet-induced animal models, genetic animal models and in the combination of both.

DNA methylation was not consistently altered in humans with HHcy. Global DNA was not always hypomethylated, but was also found to be hypermethylated in participants with high Hcy [52,58]. Studies were not similar considering tissue types, participants and the region measured. For example, global DNA methylation was measured in leukocytes and peripheral lymphocytes of vascular disease patients. Global DNA in leukocytes was hypomethylated [57], whereas in peripheral lymphocytes it was hypermethylated [58]. Disease-specific studies are limited and tissues are not always similar. Despite the clear association observed between Hcy and global or gene-specific methylation in vascular disease patients, we could not derive strong conclusions. In case of other diseases, a few studies showed association, while others did not. Hence, there is a need for more disease-specific studies which can evaluate and compare similar regions and tissues.

Furthermore, not many large scale genome-wide studies have been published until now, except for two [56,77]. In patients with schizophrenia and HHcy, no change in leukocytic global DNA methylation was observed [75], but genome-wide changes in DNA methylation were found [77]. Hence, more genome-wide studies should also be done to evaluate more site-specific genome regions and genes involved in Hcy-induced DNA methylation alteration.

Moreover, variable tissue types must be measured for DNA methylation and its markers to explore Hcy effects tissue-specifically. Two studies conducted in a heterozygous Cbs knockout mice observed tissue-specific effects in the H19 DMD of the liver, brain and aorta when given a HM-LF diet to induce HHcy [45,46]. Looking at all the studies, we did not notice any strongest effects of homocysteine metabolism on methylation in the liver tissue, as compared to other tissues. Although this might be expected, since plasma Hcy reflects liver stores primarily. In human studies, there is limited availability of the variable tissue data, which is clearly a major obstacle and may contribute to the lack of association between plasma Hcy and accessible tissues such as

blood. In most studies, leukocytes are used as proxies for other tissues as they are easily accessible. However, it remains questionable whether DNA methylation in leukocytes is representative of what is occurring in tissues. A few recent studies have investigated this possible correlation between DNA methylation in blood and other tissues. For example, leukocytic DNA methylation of selected imprinted genes like insulin-like growth factor 2 (IGF2) may serve as a proxy for mammary tissue DNA methylation in women with or without breast cancer [99]. In addition, leukocytic DNA methylation of LINE-1 repetitive elements may also serve as a proxy showing its correlation with that of colon tissue [100]. Similar studies comparing leukocytic DNA with tissue DNA, but in relation to Hcy metabolism should be performed to account for tissue-specificity.

In both animal and human studies, SAM:SAH ratio was usually observed to be lowered in HH mice or participants, in comparison to controls. In animals, not many studies were able to show strong tissue-specificity, except for one [43], where SAM:SAH ratio was increased in the liver and lowered in the heart of HH Mtrr (+/−) and Mtrr (−/−) mice as compared to Mtrr (+/+) mice. Tissue-specificity on the relationship of Hcy concentrations and SAM:SAH ratio was also observed in few studies showing lowered SAM:SAH ratio in the liver, but no change in the brain and colon of HH mice and rats, as compared to their controls [15,31,32,35,46]. As mentioned before, since plasma Hcy reflects liver stores primarily, it might suggest that this relationship could be the strongest in the liver, as compared to other tissues.

In Hcy-lowering trials, animal studies showed contrasting results in the liver, colon or brain tissues. Some studies restored methylation patterns, while others did not. But as mentioned before, this might be due to different pathophysiology between models. In humans, Hcy-lowering trials did not result in a clear improvement of DNA methylation patterns in most studies. All the reviewed trials were able to restore Hcy concentrations, but this only restored DNA methylation levels in 2 out of 10 studies [71,84]. In few studies, despite the longer trials of 1, 2 or 3 years, the methylation patterns did not change.

ACKNOWLEDGMENTS

We gratefully thank Prof. Jan Lindemans (Department of Clinical Chemistry, Erasmus University Medical Center, Rotterdam, the Netherlands) for his valuable comments on the manuscript.

SUPPLEMENTARY MATERIAL

Table 1. Diet-induced HHcy and DNA methylation in animals

Diet given[A] (v/s. Control)	Treatment period (weeks)	Hcy[B] (↑) (Plasma/Serum)	SAM (Tissue)	SAH (Tissue)	SAM:SAH ratio (Tissue)	DNA Methylation (Tissue)	Genome region	Reference
HIGH METHIONINE DIET								
HM (Cbs +/+ mice)	7 or 15	Mild	↓ (Brain) [At 7 wks], NS (Liver)	↑ (Liver) [At 15 wks], NS (Brain)	↓ (Liver, Brain) [At 15 wks]	NQ	NQ	Dayal S, 2001
HM (Sprague-Dawley rats)	4	Intermediate	↓ (Aorta)	↑ (Aorta)	↓ (Aorta)	↓ (Aorta)	B1 elements	Jiang Y, 2007
HM (ApoE -/- mice)	15	Moderate	↑ (Aorta)	↑ (Aorta)	↑ (Aorta)	↓ (Aorta)	Global, B1 elements	Jiang Y, 2012
HM (ApoE -/- mice)	8	NQ	NQ	↑ (Plasma)	NQ	↓ (Aorta)	Global	Liu C, 2008
HM (ApoE -/- mice)	15	Moderate	↓ (Cardiac)	NS (Cardiac)	NS (Cardiac)	↓ (Cardiac)	*Trp53*	Ma S, 2013
HM (Wistar rats)	20	Severe	↑ (Plasma, Bone)	↑ (Plasma, Bone)	↓ (Plasma, Bone)	NQ	NQ	Herrmann M, 2009
HM (Wistar rats)	~22	Severe	↑ (Plasma, Brainstem), NS (Frontal cortex)	↑ (Frontal cortex, Plasma, Brainstem)	↓ (Frontal cortex, Plasma, Brainstem)	NQ	NQ	Obeid R, 2011
HM (Mthfr +/+ mice)	7-15	Mild [NS]	↓ (Brain), NS (Liver)	↑ (Liver, Brain)	↓ (Liver, Brain)	NS (Liver, Brain)	Global	Devlin AM, 2004
HM (C57BL/6J mice)	4 or 8	Moderate	NQ	NQ	NQ	↓ (Circulating leucocytes)	CTCF binding sites on *mTERT* promoter	Zhang D, 2013
B-VITAMIN DEFICIENT (FOLATE, B6, B12) DIETS								
FD-MD (C57BL/6J mice)	24	Mild	↓ (Liver, Brain, Kidney, Testes)	↑ (Liver), NS (Brain, Kidney, Testes)	↓ (Liver, Brain, Kidney, Testes)	NS (Liver, Brain, Kidney, Testes)	Global	Caudill MA, 2001
LF-LB6-LB12 (ApoE -/- mice)	8	NQ	NQ	NS (Plasma)	NQ	NS(Aorta)	Global	Liu C, 2008
HM-LF-LB6-LB12 (ApoE -/- mice)	8	NQ	NQ	↑(Plasma)	NQ	↓ (Aorta)	Global	Liu C, 2008
FD or FD-B12D (Wistar rats)	8	NQ	↓ (Liver)	NS (Liver)	↓ (Liver)	NQ	NQ	Hirsch S, 2008
FD or FD-B6D-B12D	16	Moderate	↓ (Liver)	↑ (Liver)	↓ (Liver)	NS (Liver, Aorta,	Global	McNeil CJ,

Diet given[A] (v/s. Control)	Treatment period (weeks)	Hcy[B] (↑) (Plasma/Serum)	SAM (Tissue)	SAH (Tissue)	SAM:SAH ratio (Tissue)	DNA Methylation (Tissue)	Genome region	Reference
(ApoE -/- mice)						Tunica adventitia, Heart,)		2011, 2012
FD-B12D-B2D (Wistar rats)	4 (before pregnancy until weaning of Offspring) [Mothers]	Moderate [Mothers], Mild [Offspring]	↓[D] [Offspring]	NS[D] [Offspring]	↓[D] [Offspring]	NQ	NQ	Blaise S, 2005
FD (Rowett Hooded rats)	2 (prior to mating until Day 21 of gestation)	NQ (Maternal serum)	NQ	NQ	NQ	NS (Maternal & fetal livers)	Global	Maloney CA,2007
FD[E] (Sprague-Dawley rats)	4 (until mating) [Parents]	Mild/Moderate [Offspring]	NQ	NQ	NQ	↓ (Liver) [Offspring]	Global	Mejos KK, 2013
FD (Mthfr +/+BALB/c mice)	10-12 [Mothers]	Intermediate [Mothers]	↓ (Liver) [Mothers], NS (Placenta) [Embryos]	↑ (Liver) [Mothers] ↑ (Placenta) [Embryos]	↓ (Liver) [Mothers] ↓ (Placenta) [Embryos]	NQ	NQ	Mikael LG, 2013
FD (Sprague-Dawley rats)	4.3	Mild	NS (Brain)	NS (Brain)	NS (Brain)	↓ (Brain)	Global	Partearroyo T, 2013
FD-MD (Fisher 344 rats)	36	NQ	↓ (Liver), NS (Brain)	NS (Liver, Brain)	↓ (Liver), NS (Brain)	↓ (Liver), ↑ (Brain)	Global	Pogribny IP, 2006, 2008
HM-LF (Cbs +/+ C57BL/6J mice)	~35	Mild	NQ	↑ (Brain)	↓ (Brain)	NQ	NQ	Sontag E, 2007
FD (Sprague-Dawley rats)	5	NQ	NS (Colonic mucosa)	↑ (Colonic mucosa)	↓ (Colonic mucosa)	↑ (Colon) [At 3 wks]	Global	Sohn KJ, 2003
FD (Hooded-Lister (Rowett strain)rats)	24	NQ	NS (Liver, Colon)	NS (Liver, Colon)	↓ (Liver), NS (Colon)	NS (Liver, Colon)	Global	Duthie SJ, 2010
LF (Mthfr +/+ mice)	7-15	Mild	↓ (Liver), NS (Brain)	↑ (Brain), NS (Liver)	↓ (Liver, Brain)	NS (Liver, Brain)	Global	Devlin AM, 2004
HM-LF (Mthfr +/+ mice)	7-15	Mild	↓ (Liver), NS (Brain)	↑ (Brain), NS (Liver)	↓ (Liver, Brain)	NS (Liver, Brain)	Global	Devlin AM, 2004

[A] Diet given: HM = High Methionine; LF = Low folate; FD = Folate deficient; MD = Methionine deficient; B6D = Vitamin B6 deficient; B12D = Vitamin B12 deficient; LB6 = Low vitamin B6; LB12 = Low vitamin B12, [B] Hcy concentrations: Mild = <15 µmol/L; Moderate = between 15 & 30 µmol/L; Intermediate = between 30 & 100 µmol/L; Severe = >100 µmol/L, [C] NS = No significant change, NQ = Not quantified, [D] Tissue not mentioned,
[E] Folate deficient (v/s. Folate supplemented) diet

Table 2. Genetically-induced HHcy and DNA methylation in animals

Mice (v/s. non-knockout mice)	Hcy[A] (↑) (Plasma/ Serum)	SAM (Tissue)	SAH (Tissue)	SAM:SAH ratio (Tissue)	DNA Methylation (Tissue)	Genome region	Reference
Mthfr +/- mice	Mild	↓ (Testes, Ovaries), NS (Liver, Brain)	↑ (Liver, Brain), NS(Testes, Ovaries)	NQ	↓ (Brain, Ovaries), NS (Liver, Testes)	Global	Chen Z, 2001
	Mild	↓ (Liver), NS (Brain)	↑ (Brain), NS (Liver)	↓ (Liver, Brain)	NS (Liver, Brain)	Global	Devlin AM, 2004
Mthfr -/- mice	Intermediate	↓(Brain, Testes, Ovaries), NS (Liver)	↑ (Liver, Brain, Testes, Ovaries)	NQ	↓ (Brain, Ovaries), NS (Liver, Testes)	Global	Chen Z, 2001
	NQ	NQ	NQ	NQ	↓ (Cerebellum)	Global	Jadavji NM, 2012
	NQ	NQ	NQ	NQ	↓ (Hippocampus)	*Nr3c1*exon region	Jadavji NM, 2012
Cbs +/- mice	Mild	↓ (Liver, Kidney, Brain, Testes)	↑ (Liver, Testes), NS (Kidney, Brain)	↓ (Liver, Kidney, Brain, Testes)	↓ (Testes), NS (Liver, Kidney, Brain)	Global	Caudill MA, 2001
	Moderate	↑ (Brain), NS (Liver)	↑ (Liver, Brain)	↓ (Liver, Brain)	NQ	NQ	Dayal S, 2001
	Mild	NQ	NQ	NQ	↓ (Brain)	*Ppp2r1a*	Sontag E, 2007
Cbs -/- mice	Severe	↑ (Liver), NS (Kidney, Brain)	↑ (Liver, Kidney, Brain)	↓ (Liver, Kidney, Brain)	↓ (Liver), NS (Kidney, Brain)	Global	Choumenkovitch SF, 2002
	NQ[B]	↑ (Liver)	↑ (Liver)	NQ	NQ	NQ	Alberto JM, 2007
	Severe	↑ (Liver)	↑ (Liver)	↓ (Liver)	NQ	NQ	Ikeda K, 2011
Mtr +/- mice	NS	NS (Liver, Brain)	NS (Liver, Brain)	NS (Liver, Brain)	NQ	NQ	Dayal S, 2005
Mtrr +/- mice	Mild	↑ (Liver)	↓ (Liver), ↑ (Heart)	↑ (Liver), ↓ (Heart)	NQ	NQ	Elmore CL, 2007
Mtrr -/- mice	Moderate	↑ (Liver)	↓ (Liver), ↑ (Heart)	↑ (Liver), ↓ (Heart)	NQ	NQ	Elmore CL, 2007
	Moderate [Mothers]	NQ	NQ	NQ	↓ (Liver, Uterus) [Mothers], ↓ (Placenta) [Wild-type grandprogeny] Majority ↑ (Placenta) [Wild-type grandprogeny]	Global Imprinted loci (45-70% of DMRs)	Padmanabhan N, 2013

[A] Hcy concentrations: Mild = <15 µmol/L; Moderate = between 15 & 30 µmol/L; Intermediate = between 30 & 100 µmol/L; Severe = >100 µmol/L

[B] Hcy measured in the liver tissue, [C] NS = No significant change, NQ = Not quantified

Table 3. Genetically- and diet- induced HHcy in animals

Mice (v/s. non-knockout mice)	Diet[A] (v/s. Control)	Treatment period (weeks)	Hcy[B] (↑) (Plasma/Serum)	SAM (Tissue)	SAH (Tissue)	SAM / SAH (Tissue)	DNA Methylation (Tissue)	Genome region	Ref.
Cbs +/- mice	MD or FD-MD	24	Moderate	↓ (Liver, Kidney, Brain, Testes)	↑ (Liver, Testes, Brain), NS (Kidney)	↓ (Liver, Kidney, Brain, Testes)	↓ (Testes, Brain), NS (Liver, Kidney)	Global	Caudill MA, 2001
	HM-LF	~39-52	Intermediate	NS (Liver, Brain)	↑ (Liver, Brain)	↓ (Liver, Brain)	↓ (Liver), ↑ (Brain), ↑ (Aorta)	H19 DMD	Devlin AM, 2005
	HM-LF	3	Mild	NS (Liver, Brain)	↑ (Liver), NS (Brain)	↓ (Liver), NS (Brain)	↓ (Liver), ↑ (Brain)	Maternal H19 DMD	Glier MB, 2013
	HM-FR	7	Moderate	↑ (Brain), NS (Liver)	↑ (Liver, Brain)	↓ (Liver), NS (Brain)	NQ	NQ	Dayal S, 2001
		15	Moderate	NS (Brain, Liver)	↑ (Liver, Brain)	↓ (Liver, Brain)			
	HM-LF	12	Moderate	NQ	NQ	NQ	↑ (Aorta)	Nr3c1 [promoter transcripts]	Sulistyoningrum DC, 2012
	HM-LF	3-9	Intermediate	NS (Liver)	↑ (Liver)	↓ (Liver)	↑ (Liver)	Fads2 promoter	Devlin AM, 2007
	HM-LF	3-12	Intermediate	NQ	NQ	NQ	↓ [Soat2] (Liver) ↑ [B1] (Liver)	5' portion of the Soat2, B1 repetitive elements	Devlin AM, 2010

[A] Diet given: HM = High Methionine; LF = Low folate; FD = Folate deficient; FR = Folate replete; MD = Methionine deficient

[B] Hcy concentrations: Mild = <15 µmol/L; Moderate = between 15 & 30 µmol/L; Intermediate = between 30 & 100 µmol/L; Severe = >100 µmol/L

[C] NS = No significant change, NQ = Not quantified

Table 4. HHcy and DNA methylation in humans

Participants	Plasma Hcy[A] (↑)	SAM	SAH	SAM:SAH ratio	DNAmethylation (Tissue)	Region	Reference
Healthy Vegetarians (n=71) v/s. Omnivores (n=79)	Mild	NQ	NQ	NQ	NQ (Peripheral blood)[U]	Global (Alu repeat elements); 3 CpG sites in the p66shc promoter	Geisel J, 2005
Healthy Adult Females	Mild	NS (Plasma)	↑ (Plasma, Lymphocytes)	↓ (Plasma)	↓ (Lymphocytes)[N]	Global	Melnyk S, 2000; Yi P, 2000
Healthy Middle-aged Adults	Mild	NQ	NQ	NQ	NQ (Peripheral leukocytes)[P]	LINE-1	Perng W, 2014
High (n=25) v/s. Low (n=25) Hcy samples	Moderate/ Intermediate/ Severe	NQ	NQ	NQ	↓ [14 CpGs][N], ↑ [3 CpGs][P] (Leukocytes)	RFC1 gene: 17 CpGs	Farkas SA, 2013
PREGNANCY							
Pregnant Women (n=24)	Mild (Fetal Cord)	NQ	NQ	NQ	↓ (Cord blood)[N]	Global (LINE-1 elements)	Fryer AA, 2009
Fetal cord blood samples (n=12)	Mild (Cord)	NQ	NQ	NQ	↑↓ (Cord blood)[P,N]	Genome-wide CpGs: 298 autosomal, 8 X chromosome	Fryer AA, 2011
VASCULAR DISORDERS							
Atherosclerotic patients (n=17) v/s. controls (n=15)	Mild	NS (Plasma)	↑ (Plasma)	↓ (Plasma)	↓ (Leukocytes)[N]	Global	Castro R, 2003
CAD patients (n=137) v/s. controls (n=150)	Mild/ Moderate/ Intermediate	NQ	NQ	NQ	↑ (Peripheral lymphocytes)[P]	Global	Sharma P, 2008
Elderly men (n=712) [No Ischemic heart disease or stroke (n=470), Ischemic heart disease (n=212), stroke (n=51), Ischemic heart disease or stroke (n=242)]	Mild [NS]	NQ	NQ	NQ	↓ (Blood)[U]	LINE-1	Baccarelli A, 2010
Atheromatosis patients (n=54) v/s. controls (n=28)	Moderate	NQ	NQ	NQ	↑ (Blood)[P]	ESR1 promoter	Huang YS, 2009
Atherosclerosis patients (n=197) v/s. controls (n=165)	Moderate (Serum)	NQ	NQ	NQ	↓ (Leukocytes)[N]	hTERT promoter	Zhang D, 2013
CAD patients	Mild/ Moderate	NQ	NQ	NQ	↓ (Peripheral blood leukocytes)[N]	SHC1 promoter region	Kim CS, 2011

Participants	Plasma Hcy[A] (↑)	SAM	SAH	SAM:SAH ratio	DNAmethylation (Tissue)	Region	Reference
Patients who underwent coronary artery bypass grafting (n=192)	Moderate	NQ	NQ	NQ	↑ (Internal mammary artery)[P], ↓ (Punch aortic fragments)[N]	NOS2 proximal promoter	Rodriguez-Esparragon F, 2012
Patients with CBS deficiency (Treated with Hcy-lowering therapy) v/s. Controls)	Intermediate	↑ (Plasma)	↑ (Plasma)	NS (Plasma)	NS (Heparin blood)[U]	Global, H19 DMR	Heil SG, 2007
CANCER							
Colorectal neoplasia patients (cancer, n=28; adenoma, n=35) v/s. controls (n=76)	Mild	NQ	NQ	NQ	↓ (Normal-appearing colonic mucosa) (Leukocytes)[N]	Global	Pufulete M, 2003
Colorectal neoplasia patients (cancer, n=28; adenoma, n=35; hyperplastic polyps, n=17) v/s. controls (n=76)	Mild	NQ	NQ	NQ	↑ (Normal-appearing colorectal mucosa)[P]	ESR1 promoter region	Al-Ghnaniem R, 2007
Colorectal cancer cases (n=190) v/s. Double, matched referents	Mild [NS]	NQ	NQ	NQ	NS (Colorectal tumor)[U]	CDKN2A, MLH1, CACNA1G, NEUROG1, RUNX3, SOCS1, IGF2 & CRABP1 promoter	Van Guelpen B, 2010
Esophageal disease patients (including cancer) (n=124) v/s. controls	≥ Moderate	NQ	NQ	NQ	↑ (Whole blood)[P]	MLH1 promoter	Vasavi M, 2006
Primary breast cancer patients (n=137)	Mild	NQ	NQ	NQ	↑ (Breast tumor)[P]	RARB promoter exon, ESR1 exon 1 regions	Pirouzpanah S, 2010
Specimens from breast carcinomas (n = 40) & peripheral blood samples (n = 139) from breast cancer patients	NQ	NQ	NQ	NQ	↑ (Breast cancer tissue & corresponding peripheral blood)[P]	BRCA1 promoter	Naushad SM, 2014
					↑ (Breast cancer tissue)[P]	RASSF1 promoter (ER/PR-negative breast cancers)	
RENAL DISEASES							
Uremia patients (n=32) v/s. controls (n=11)	Intermediate	NQ	NQ	NQ	↓ (PBMCs)[N]	Global	Ingrosso D, 2003
Stage 2-4 chronic kidney disease patients (n=78)	Moderate	NQ	NQ	NQ	NQ (Leukocytes)[U]	Global	Nanayakkara PW, 2008
Chronic hemodialysis patients (n=20) v/s. Normal subjects (n=20)	Moderate (Serum)	NQ	NQ	NQ	NS (Peripheral blood leukocytes)[U]	Global	Hsu CY, 2012

Participants		Plasma Hcy[A] (↑)	SAM	SAH	SAM:SAH ratio	DNAmethylation (Tissue)	Region	Reference
BRAIN DISORDERS								
Alzheimer's patients (n=26) v/s. controls (n=29)		Mild/ Moderate	↑ (Plasma)	↑ (Plasma)	↓ (Plasma)	NQ	NQ	Selley ML, 2007
Euthymic bipolar patients (n=40) v/s. Healthy controls (n=27)		Mild	NQ	NQ	NQ	NS (Peripheral blood leukocytes)[U]	Global	Bromberg A, 2009
Schizophrenia patients (n=28) v/s. Healthy controls (n=26)		Mild/ Moderate	NQ	NQ	NQ	NS (Peripheral blood leukocytes)[U]	Global	Bromberg A, 2008
Schizophrenia patients (n=42) v/s. controls (n=42)		Moderate	NQ	NQ	NQ	↑↓ (Leukocytes)[P,N]	1,338 CpGs [Genome-wide]	Kinoshita M, 2013
GENETIC VARIANTS								
Subjects with Mthfr 677 TT v/s. Mthfr 677 CC		Moderate	NQ	NQ	NQ	↓ (PBMCs)[N]	Global	Friso S, 2002
Subjects with Mthfr 677TT		Mild	NQ	NQ	NQ	↓ (Leukocytes)[N]	Global	Castro R, 2004
Subjects with Mthfr 677 TT v/s. Mthfr 677 CC	Week 7 (Folate post-depletion)	Mild	NQ	NQ	NQ	↓ (Leukocytes)[N]	Global	Shelnutt KP, 2004
	Week 14 (Folate post-repletion)	Mild				↑ (Leukocytes)[P] [In subjects with Mthfr 677 TT only]		
Subjects with Mthfr 677 TT v/s. Mthfr 677 CC	Week 7 (Folate post-depletion)	Mild	NQ	NQ	NQ	NS (Leukocytes)[U]	Global	Axume J, 2007
	Week 14 (Folate post-repletion)	Mild				↓ (Leukocytes)[N]		
Subjects with Mthfr 677 TT v/s. Mthfr 677 CC		Mild	NQ	NQ	NQ	NS (Lymphocytes)[U]	Global	Narayanan S, 2004
Subjects with Mthfr 677 TT v/s. Mthfr 677 CC		Mild [NS]	NQ	NQ	NQ	NS (Colon)[U]	Global	Hanks J, 2013
Subjects with Mthfr 677 TT v/s. Mthfr 677 CC		Moderate	NQ	NQ	NQ	NS (Colonic mucosa)[U]	Global	Pufulete M, 2005

[A] Hcy concentrations: Mild = <15 µmol/L; Moderate = between 15 & 30 µmol/L; Intermediate = between 30 & 100 µmol/L; Severe = >100 µmol/L

[N] Negative correlation with Hcy, [P] Positive correlation with Hcy, [U] Undefined trend of correlation

[C] NS = No significant change, NQ = Not quantified

Table 5. Effects of Hcy-lowering therapy in Animals

Animal Type	Amount (Folate)	Treatment period (weeks)	Hcy (Plasma)	SAM (Tissue)	SAH (Tissue)	SAM:SAH ratio (Tissue)	DNA methylation (Tissue)	Genome region	Reference
Rats	40 mg FA/kg diet	4	↓	NS (Liver)	NS (Liver)	NS (Liver)	NS (Liver)	Global	Achon M, 2007
Rats (under B12D diet)	100 mg FA/kg diet	6	↓	NS (Liver, Brain)	NS (Liver, Brain)	NS (Liver, Brain)	NS (Liver)	Global	Min H, 2009
Rat offsprings [At weaning]	5 mg FA/kg diet [Maternal]	3*	↓	NQ	NQ	NQ	↓ (Liver) [Global, *Pparg*, *Esr1*, *Trp53*, *Apc*]	Global	Sie KK, 2013
Rat offsprings [At Post-weaning]	NA [Post-weaning]	11	↓	NQ	NQ	NQ	↓ (Liver) [Global][C] ↑ (Liver) [*Pparg*, *Trp53*, *p16*][B,C] [*Esr1*, *Apc*][C]	*Pparg* promoter; *Esr1* promoter; *Trp53* exon region; *Apc* exon region; *p16* promoter	
Rats	8 mg FA/kg diet	5	↓	NS (Colonic mucosa)	NS (Colonic mucosa)	NS (Colonic mucosa)	NS (Colon)	Global; *Trp53* promoter & exon region	Sohn KJ, 2003
Rat offsprings [At weaning]	5 mg FA/kg diet [Maternal]	3*	↓	NQ	NQ	NQ	↑ (Colon)[B]	Global	Sie KK, 2011
Rat offsprings [At Post-weaning]	NA [Post-weaning]	11	↓	NQ	NQ	NQ	↓ (Colon)[B,D]		
Rats	8 or 40 mg FA/kg	~4	NS	NS (Brain)	NS (Brain)	NS (Brain)	NS (Brain)	Global	Partearroyo T, 2013
CBS +/- mice (under HM diet)	0.0057 µg/g bodyweight per day FA in drinking water	4	↓	NQ	NQ	NQ	NS (Brain)	Global	Kalani A, 2014

[A] NS = No significant change; FA = Folic acid, NA = Not applicable, NQ = Not Quantified
[B] Effect due to maternal supplement alone
[C] Effect due to post-weaning supplementation alone, [D] Effect due to both maternal and post-weaning supplementation
*During pregnancy & lactation

Table 6. Effects of Hcy-lowering therapy in humans

Participants	Dietary supplement	Amount	Treatment period	Plasma Hcy	DNA Methylation (Tissue)	Region	Reference
Colorectal adenoma patients	Folate	400 µg/day (FA)	10 weeks	↓	↑ (Leukocytes)[b]	Global	Pufulete M, 2005
Colorectal adenoma patients	Folate	400 µg/day (FA)	10 weeks	↓	NS (Colonic mucosa)	CpGs of *ESR1*&*MLH1* genes	Al-Ghnaniem Abbadi RP, 2012
Patients with previous colorectal adenoma	Folate & vitamin B12	5 mg/day (FA), 1.25 mg/day (Vit B12)	6 months	↓	NS (Rectal mucosa)	Promoter methylation of 6 tumor suppressor & DNA repair genes	van den Donk M, 2007
Patients with HHcy & uremia	Folate	15 mg/day (oral methyltetrahydrofolate)	2 months	↓	↑ (PBMCs)[b]	Global	Ingrosso D, 2003
Participants with intermediate HHcy	Folate	5 mg/day (FA)	8 weeks	↓	NS(PBMCs)	Global	Pizzolo F, 2011
Long-term stable dialysis patients	Folate (or) Folate & vitamin B12	15 mg thrice weekly (FA); 1000 µg once a week (Vit B12)	12-20 weeks	↓	NS(Lymphocyte)	Global	Stopper H, 2008
Stage 2-4 chronic kidney disease patients	B-vitamins	5 mg/d (FA), 100 mg/d (pyridoxine), 1 mg/d (Vit B12)	1 year	NQ	NS (Leukocytes)	Global	Nanayakkara PW, 2008
Adults	B & D vitamins	500 µg/d (FA), 500 µg (Vit B12), 50 mg (Vit B6), 1200 IU (Vit D), 456 mg (calcium)	1 year	↓	NS (Whole blood)	LINE-1	Hubner U, 2013
Older healthy group with elevated Hcy of >13 µmol/l	Folate & B vitamins	1 mg/d (folate), 500 µg/d (Vit B12), 10 mg (Vit B6)	2 years	↓	NS (Plasma)[c]	NQ	Green TJ, 2010
Participants with moderate HHcy	Folate	0.8 mg/day (FA)	3 years	↓	NS (Peripheral blood leukocytes)	Global	Jung AY, 2011
Postmenopausal women	Folate fortification	NA	2 years	↓	NS (Leukocytes)	Global	Bae S, 2014

[A] NQ = Not quantified; NS = No significant change; FA = Folic acid, NA = Not applicable
[B] Hcy-lowering trial significantly restored DNA methylation patterns, [C] Measured SAM, SAH and SAM:SAH ratio

REFERENCES

1. J. Selhub, Homocysteine metabolism, Annu. Rev. Nutr. 19 (1999) 217–246.
2. C.J. Boushey, et al., A quantitative assessment of plasma homocysteine as a risk factor for vascular disease. Probable benefits of increasing folic acid intakes, JAMA 274 (13) (1995) 1049–1057.
3. J.B. van Meurs, et al., Homocysteine levels and the risk of osteoporotic fracture, N. Engl. J. Med. 350 (20) (2004) 2033–2041.
4. W. Fu, et al., Interrelations between plasma homocysteine and intracellular S-adenosylhomocysteine, Biochem. Biophys. Res. Commun. 271 (1) (2000) 47–53.
5. S.J. James, et al., Elevation in S-adenosylhomocysteine and DNA hypomethylation: potential epigenetic mechanism for homocysteine-related pathology, J. Nutr. 132 (8 Suppl) (2002) 2361S–2366S.
6. J. Selhub, et al., Vitamin status and intake as primary determinants of homocysteinemia in an elderly population, JAMA 270 (22) (1993) 2693–2698.
7. N. Mahalle, et al., Vitamin B12 deficiency and hyperhomocysteinemia as correlates of cardiovascular risk factors in Indian subjects with coronary artery disease, J. Cardiol. 61 (4) (2013) 289–294.
8. N.E. Bergen, et al., Homocysteine and folate concentrations in early pregnancy and the risk of adverse pregnancy outcomes: the Generation R Study, BJOG 119 (6) (2012) 739–751.
9. T. Zhang, et al., Maternal serum vitamin B12, folate and homocysteine and the risk of neural tube defects in the offspring in a high-risk area of China, Public Health Nutr. 12 (5) (2009) 680–686.
10. K.A. da Costa, et al., Choline deficiency in mice and humans is associated with increased plasma homocysteine concentration after a methionine load, Am. J. Clin. Nutr. 81 (2) (2005) 440–444.
11. S.E. Chiuve, et al., The association between betaine and choline intakes and the plasma concentrations of homocysteine in women, Am. J. Clin. Nutr. 86 (4) (2007) 1073–1081.
12. J. Chan, et al., Low dietary choline and low dietary riboflavin during pregnancy influence reproductive outcomes and heart development in mice, Am. J. Clin. Nutr. 91 (4) (2010) 1035–1043.
13. J.B. van Meurs, et al., Common genetic loci influencing plasma homocysteine concentrations and their effect on risk of coronary artery disease, Am. J. Clin. Nutr. 98 (3) (2013) 668–676.
14. S.S. Kang, P.W. Wong, M.R. Malinow, Hyperhomocyst(e)inemia as a risk factor for occlusive vascular disease, Annu. Rev. Nutr. 12 (1992) 279–298.
15. S. Dayal, et al., Endothelial dysfunction and elevation of S-adenosylhomocysteine in cystathionine beta-synthase-deficient mice, Circ. Res. 88 (11) (2001) 1203–1209.
16. Y. Jiang, et al., Hyperhomocysteinemia-mediated DNA hypomethylation and its potential epigenetic role in rats, Acta Biochim. Biophys. Sin. 39 (9) (2007) 657–667.
17. Y. Jiang, et al., The comprehensive effects of hyperlipidemia and hyperhomocysteinemia on pathogenesis of atherosclerosis and DNA hypomethylation in ApoE−/− mice, Acta Biochim. Biophys. Sin. 44 (10) (2012) 866–875.
18. C. Liu, et al., Plasma S-adenosylhomocysteine is a better biomarker of atherosclerosis than homocysteine in apolipoprotein E-deficient mice fed high dietarymethionine, J. Nutr. 138 (2) (2008) 311–315.
19. S. Ma, et al., Hyperhomocysteinemia induces cardiac injury by up-regulation of p53-dependent Noxa and Bax expression through the p53 DNA methylation in ApoE(−/−) mice, Acta Biochim. Biophys. Sin. 45 (5) (2013) 391–400.

20. M. Herrmann, et al., Hyperhomocysteinemia induces a tissue specific accumulation of homocysteine in bone by collagen binding and adversely affects bone, Bone 44 (3) (2009) 467–475.
21. R. Obeid, et al., Folate is related to phosphorylated neurofilament-H and P-tau (Ser396) in rat brain, J. Neurochem. 117 (6) (2011) 1047–1054.
22. M.A. Caudill, et al., Intracellular S-adenosylhomocysteine concentrations predict global DNA hypomethylation in tissues of methyl-deficient cystathionine beta-synthase heterozygous mice, J. Nutr. 131 (11) (2001) 2811–2818.
23. S. Hirsch, et al., Lack of effect of diet-induced hypomethylation on endotheliumdependent relaxation in rats, Clin. Nutr. 27 (6) (2008) 895–899.
24. C.J.McNeil, et al., Differential effects of nutritional folic acid deficiency andmoderate hyperhomocysteinemia on aortic plaque formation and genome-wide DNAmethylation in vascular tissue from ApoE−/− mice, Clin. Epigen. 2 (2) (2011) 361–368.
25. C.J. McNeil, et al., Nutritional B vitamin deficiency disrupts lipid metabolism causing accumulation of proatherogenic lipoproteins in the aorta adventitia of ApoE null mice, Mol. Nutr. Food Res. 56 (7) (2012) 1122–1130.
26. S. Blaise, et al., Mild neonatal hypoxia exacerbates the effects of vitamin-deficient diet on homocysteine metabolism in rats, Pediatr. Res. 57 (6) (2005) 777–782.
27. C.A. Maloney, S.M. Hay, W.D. Rees, Folate deficiency during pregnancy impacts on methyl metabolism without affecting global DNA methylation in the rat fetus, Br. J. Nutr. 97 (6) (2007) 1090–1098.
28. K.K. Mejos, et al., Effects of parental folate deficiency on the folate content, global DNA methylation, and expressions of FRalpha, IGF-2 and IGF-1R in the postnatal rat liver, Nutr. Res. Pract. 7 (4) (2013) 281–286.
29. L.G. Mikael, et al., Low dietary folate and methylenetetrahydrofolate reductase deficiency may lead to pregnancy complications through modulation of ApoAI and IFN-gamma in spleen and placenta, and through reduction of methylation potential, Mol. Nutr. Food Res. 57 (4) (2013) 661–670.
30. T. Partearroyo, et al., Dietary folic acid intake differentially affects methionine metabolism markers and hippocampus morphology in aged rats, Eur. J. Nutr. 52 (3) (2013) 1157–1167.
31. I.P. Pogribny, et al., Epigenetic alterations in the brains of Fisher 344 rats induced by long-term administration of folate/methyl-deficient diet, Brain Res. 1237 (2008) 25–34.
32. I.P. Pogribny, et al., Irreversible global DNA hypomethylation as a key step in hepatocarcinogenesis induced by dietary methyl deficiency, Mutat. Res. 593 (1–2) (2006) 80–87.
33. E. Sontag, et al., Protein phosphatase 2A methyltransferase links homocysteine metabolism with tau and amyloid precursor protein regulation, J. Neurosci. 27 (11) (2007) 2751–2759.
34. K.J. Sohn, et al., The effect of dietary folate on genomic and p53-specific DNA methylation in rat colon, Carcinogenesis 24 (1) (2003) 81–90.
35. S.J. Duthie, et al., Folate deficiency alters hepatic and colon MGMT and OGG-1 DNA repair protein expression in rats but has no effect on genome-wide DNA methylation, Cancer Prev. Res. 3 (1) (2010) 92–100.
36. A.M. Devlin, et al., Effect of Mthfr genotype on diet-induced hyperhomocysteinemia and vascular function in mice, Blood 103 (7) (2004) 2624–2629.
37. Z. Chen, et al., Mice deficient in methylenetetrahydrofolate reductase exhibit hyperhomocysteinemia and decreased methylation capacity, with neuropathology and aortic lipid deposition, Hum. Mol. Genet. 10 (5) (2001) 433–443.

38. N.M. Jadavji, et al., Severe methylenetetrahydrofolate reductase deficiency in mice results in behavioral anomalies with morphological and biochemical changes in hippocampus, Mol. Genet. Metab. 106 (2) (2012) 149–159.
39. S.F. Choumenkovitch, et al., In the cystathionine beta-synthase knockout mouse, elevations in total plasma homocysteine increase tissue S-adenosylhomocysteine, but responses of S-adenosylmethionine and DNA methylation are tissue specific, J. Nutr. 132 (8) (2002) 2157–2160.
40. J.M. Alberto, et al., Mice deficient in cystathionine beta synthase display altered homocysteine remethylation pathway, Mol. Genet. Metab. 91 (4) (2007) 396–398.
41. K. Ikeda, et al., Triacylglycerol/phospholipid molecular species profiling of fatty livers and regenerated non-fatty livers in cystathionine beta-synthase-deficient mice, an animal model for homocysteinemia/homocystinuria, Anal. Bioanal. Chem. 400 (7) (2011) 1853–1863.
42. S. Dayal, et al., Cerebral vascular dysfunction in methionine synthase-deficient mice, Circulation 112 (5) (2005) 737–744.
43. C.L. Elmore, et al., Metabolic derangement of methionine and folate metabolism in mice deficient in methionine synthase reductase, Mol. Genet. Metab. 91 (1) (2007) 85–97.
44. N. Padmanabhan, et al., Mutation in folate metabolism causes epigenetic instability and transgenerational effects on development, Cell 155 (1) (2013) 81–93.
45. A.M. Devlin, et al., Tissue-specific changes in H19 methylation and expression in mice with hyperhomocysteinemia, J. Biol. Chem. 280 (27) (2005) 25506–25511.
46. M.B. Glier, et al., Tissue-specific relationship of S-adenosylhomocysteine with allele specific H19/Igf2methylation and imprinting in mice with hyperhomocysteinemia, Epigenet. 8 (1) (2013) 44–53.
47. D.C. Sulistyoningrum, R. Singh, A.M. Devlin, Epigenetic regulation of glucocorticoid receptor expression in aorta from mice with hyperhomocysteinemia, Epigenet. 7 (5) (2012) 514–521.
48. A.M. Devlin, et al., Hypermethylation of Fads2 and altered hepatic fatty acid and phospholipid metabolism in mice with hyperhomocysteinemia, J. Biol. Chem. 282 (51) (2007) 37082–37090.
49. A.M. Devlin, et al., Hepatic acyl-coenzyme a:cholesterol acyltransferase-2 expression is decreased in mice with hyperhomocysteinemia, J. Nutr. 140 (2) (2010) 231–237.
50. J. Geisel, et al., The vegetarian lifestyle and DNA methylation, Clin. Chem. Lab. Med. 43 (10) (2005) 1164–1169.
51. S. Melnyk, et al., Measurement of plasma and intracellular S-adenosylmethionine and S-adenosylhomocysteine utilizing coulometric electrochemical detection: alterations with plasma homocysteine and pyridoxal 5′-phosphate concentrations, Clin. Chem. 46 (2) (2000) 265–272.
52. P. Yi, et al., Increase in plasma homocysteine associated with parallel increases in plasma S-adenosylhomocysteine and lymphocyte DNA hypomethylation, J. Biol. Chem. 275 (38) (2000) 29318–29323.
53. W. Perng, et al., Dietary intake, plasma homocysteine, and repetitive element DNA methylation in the Multi-Ethnic Study of Atherosclerosis (MESA), Nutr. Metab. Cardiovasc. Dis. 24 (6) (2014) 614–622.
54. S.A. Farkas, et al., Epigenetic alterations in folate transport genes in placental tissue from fetuses with neural tube defects and in leukocytes from subjects with hyperhomocysteinemia, Epigenet. 8 (3) (2013) 303–316.
55. A.A. Fryer, et al., LINE-1 DNA methylation is inversely correlated with cord plasma homocysteine in man: a preliminary study, Epigenet. 4 (6) (2009) 394–398.

56. A.A. Fryer, et al., Quantitative, high-resolution epigenetic profiling of CpG loci identifies associations with cord blood plasma homocysteine and birth weight in humans, Epigenet. 6 (1) (2011) 86–94.
57. R. Castro, et al., Increased homocysteine and S-adenosylhomocysteine concentrations and DNA hypomethylation in vascular disease, Clin. Chem. 49 (8) (2003) 1292–1296.
58. P. Sharma, et al., Detection of altered global DNA methylation in coronary artery disease patients, DNA Cell Biol. 27 (7) (2008) 357–365.
59. A. Baccarelli, et al., Ischemic heart disease and stroke in relation to blood DNA methylation, Epidemiology 21 (6) (2010) 819–828.
60. Y.S. Huang, Y.F. Zhi, S.R.Wang, Hypermethylation of estrogen receptor-alpha gene in atheromatosis patients and its correlation with homocysteine, Pathophysiology 16 (4) (2009) 259–265.
61. D. Zhang, et al., Homocysteine-related hTERT DNA demethylation contributes to shortened leukocyte telomere length in atherosclerosis, Atherosclerosis 231 (1) (2013) 173–179.
62. C.S. Kim, et al., Homocysteine promotes human endothelial cell dysfunction via site-specific epigenetic regulation of p66shc, Cardiovasc. Res. 92 (3) (2011) 466–475.
63. F. Rodriguez-Esparragon, et al., Homocysteinylated protein levels in internal mammary artery (IMA) fragments and its genotype-dependence. S-homocysteine induced methylation modifications in IMA and aortic fragments, Mol. Cell. Biochem. 369 (1–2) (2012) 235–246.
64. S.G. Heil, et al., DNA methylation status is not impaired in treated cystathionine beta-synthase (CBS) deficient patients, Mol. Genet. Metab. 91 (1) (2007) 55–60.
65. M. Pufulete, et al., Folate status, genomic DNA hypomethylation, and risk of colorectal adenoma and cancer: a case control study, Gastroenterology 124 (5) (2003) 1240–1248.
66. R. Al-Ghnaniem, et al., Methylation of estrogen receptor alpha and mutL homolog 1 in normal colonic mucosa: association with folate and vitamin B-12 status in subjects with and without colorectal neoplasia, Am. J. Clin. Nutr. 86 (4) (2007) 1064–1072.
67. B. Van Guelpen, et al., One-carbon metabolism and CpG island methylator phenotype status in incident colorectal cancer: a nested case-referent study, Cancer Causes Control 21 (4) (2010) 557–566.
68. M. Vasavi, et al., DNA methylation in esophageal diseases including cancer: special reference to hMLH1 gene promoter status, Tumori 92 (2) (2006) 155–162.
69. S. Pirouzpanah, et al., The effect of modifiable potentials on hypermethylation status of retinoic acid receptor-beta2 and estrogen receptor-alpha genes in primary breast cancer, Cancer Causes Control 21 (12) (2010) 2101–2111.
70. S.M. Naushad, et al., Impact of hyperhomocysteinemia on breast cancer initiation and progression: epigenetic perspective, Cell Biochem. Biophys. 68 (2) (2014) 397–406.
71. D. Ingrosso, et al., Folate treatment and unbalanced methylation and changes of allelic expression induced by hyperhomocysteinaemia in patients with uraemia, Lancet 361 (9370) (2003) 1693–1699.
72. P.W. Nanayakkara, et al., Association between global leukocyte DNA methylation, renal function, carotid intima–media thickness and plasma homocysteine in patients with stage 2–4 chronic kidney disease, Nephrol. Dial. Transplant. 23 (8) (2008) 2586–2592.
73. C.Y. Hsu, et al., Global DNA methylation not increased in chronic hemodialysis patients: a case–control study, Ren. Fail. 34 (10) (2012) 1195–1199.
74. M.L. Selley, A metabolic link between S-adenosylhomocysteine and polyunsaturated fatty acid metabolism in Alzheimer's disease, Neurobiol. Aging 28 (12) (2007) 1834–1839.

75. A. Bromberg, et al., No association between global leukocyte DNA methylation and homocysteine levels in schizophrenia patients, Schizophr. Res. 101 (1–3) (2008) 50–57.
76. A. Bromberg, et al., Global leukocyte DNA methylation is not altered in euthymic bipolar patients, J. Affect. Disord. 118 (1–3) (2009) 234–239.
77. M. Kinoshita, et al., Plasma total homocysteine is associated with DNA methylation in patients with schizophrenia, Epigenet. 8 (6) (2013) 584–590.
78. S. Friso, et al., A common mutation in the 5,10-methylenetetrahydrofolate reductase gene affects genomic DNA methylation through an interaction with folate status, Proc. Natl. Acad. Sci. U. S. A. 99 (8) (2002) 5606–5611.
79. R. Castro, et al., 5,10-methylenetetrahydrofolate reductase (MTHFR) 677C→T and 1298A→Cmutations are associated with DNA hypomethylation, J. Med. Genet. 41 (6) (2004) 454–458.
80. K.P. Shelnutt, et al., Methylenetetrahydrofolate reductase 677C→T polymorphism affects DNAmethylation in response to controlled folate intake in young women, J. Nutr. Biochem. 15 (9) (2004) 554–560.
81. J. Axume, et al., The MTHFR 677TT genotype and folate intake interact to lower global leukocyte DNA methylation in young Mexican American women, Nutr. Res. 27 (1) (2007) 1365-1317.
82. S. Narayanan, et al., Associations between two common variants C677T and A1298C in the methylenetetrahydrofolate reductase gene and measures of folate metabolism and DNA stability (strand breaks, misincorporated uracil, and DNA methylation status) in human lymphocytes in vivo, Cancer Epidemiol. Biomarkers Prev. 13 (9) (2004) 1436–1443.
83. J. Hanks, et al., The association between MTHFR 677C N T genotype and folate status and genomic and gene-specific DNA methylation in the colon of individuals without colorectal neoplasia, Am. J. Clin. Nutr. 98 (6) (2013) 1564–1574.
84. M. Pufulete, et al., Influence of folate status on genomic DNAmethylation in colonic mucosa of subjects without colorectal adenoma or cancer, Br. J. Cancer 92 (5) (2005) 838–842.
85. M. Achon, et al., Supranormal dietary folic acid supplementation: effects on methionine metabolism in weanling rats, Br. J. Nutr. 98 (3) (2007) 490–496.
86. H. Min, Effects of dietary supplementation of high-dose folic acid on biomarkers of methylating reaction in vitamin B(12)-deficient rats, Nutr. Res. Pract. 3 (2) (2009) 122–127.
87. K.K. Sie, et al., Effect of maternal and postweaning folic acid supplementation on global and gene-specific DNA methylation in the liver of the rat offspring, Mol. Nutr. Food Res. 57 (4) (2013) 677–685.
88. K.K. Sie, et al., Effect of maternal and postweaning folic acid supplementation on colorectal cancer risk in the offspring, Gut 60 (12) (2011) 1687–1694.
89. A. Kalani, et al., Nutri-epigenetics ameliorates blood–brain barrier damage and neurodegeneration in hyperhomocysteinemia: role of folic acid, J. Mol. Neurosci. 52 (2) (2014) 202–215.
90. M. Pufulete, et al., Effect of folic acid supplementation on genomic DNAmethylation in patients with colorectal adenoma, Gut 54 (5) (2005) 648–653.
91. R. Al-Ghnaniem Abbadi, P. Emery, M. Pufulete, Short-term folate supplementation in physiological doses has no effect on ESR1 and MLH1 methylation in colonic mucosa of individuals with adenoma, J. Nutrigenet. Nutrige. 5 (6) (2012) 327–338.
92. M. van den Donk, et al., Folic acid and vitamin B-12 supplementation does not favorably influence uracil incorporation and promoter methylation in rectal mucosa DNA of subjects with previous colorectal adenomas, J. Nutr. 137 (9) (2007) 2114–2120.

93. F. Pizzolo, et al., Folic acid effects on s-adenosylmethionine, s-adenosylhomocysteine, and DNA methylation in patients with intermediate hyperhomocysteinemia, J. Am. Coll. Nutr. 30 (1) (2011) 11–18.
94. H. Stopper, et al., Reduction of the genomic damage level in haemodialysis patients by folic acid and vitamin B12 supplementation, Nephrol. Dial. Transplant. 23 (10) (2008) 3272–3279.
95. U. Hubner, et al., Effect of 1 year B and D vitamin supplementation on LINE-1 repetitive element methylation in older subjects, Clin. Chem. Lab. Med. 51 (3) (2013) 649–655.
96. T.J. Green, et al., Homocysteine-lowering vitamins do not lower plasma S-adenosylhomocysteine in older people with elevated homocysteine concentrations, Br. J. Nutr. 103 (11) (2010) 1629–1634.
97. A.Y. Jung, et al., No effect of folic acid supplementation on global DNA methylation in men and women with moderately elevated homocysteine, PLoS One 6 (9) (2011) e24976.
98. S. Bae, et al., Impact of folic acid fortification on global DNA methylation and one-carbon biomarkers in the Women's Health Initiative Observational Study cohort, Epigenet. 9 (3) (2014) 396–403.
99. L. Barault, et al., Leukocyte DNA as surrogate for the evaluation of imprinted Loci methylation in mammary tissue DNA, PLoS One 8 (2) (2013) e55896.
100. J.E. Ashbury, et al., Biomarkers measured in buccal and blood leukocyte DNA as proxies for colon tissue global methylation, Int. J. Mol. Epidemiol. Genet. 5 (2) (2014) 120–124.

CHAPTER 3

Nutrients and DNA methylation across the life course: a systematic review

Kim V.E. Braun*, Pooja R. Mandaviya*, Oscar H. Franco, Jana Nano, Carolin Thomas, Wichor M. Bramer, Taulant Muka, Jenna Troup, Joyce B.J. van Meurs, Sandra G. Heil, Trudy Voortman

*These authors contributed equally as first authors

To be submitted

ABSTRACT

Background and objectives: DNA methylation can be modified by environmental factors, including nutrition. In order to gain more insight in effects of nutrients on DNA methylation, we conducted a systematic review on the relation between nutrients and DNA methylation in humans across the life course.

Methods: The literature search was designed by an experienced biomedical information specialist. Six bibliographic databases (Embase.com, Medline (Ovid), Web-of-Science, PubMed, Cochrane Central and Google Scholar) were searched. We selected studies that examined the association between nutrients (blood levels; dietary intake; or dietary supplements) and DNA methylation (global, site specific, or genome-wide) in humans of any age, with no restrictions on year of publication, language, or study design. Abstract screening, full text selection, and data extraction was performed by two independent reviewers, with a third reviewer available to solve any disagreements.

Results: We identified 3774 references, of which 98 studies met all inclusion criteria. The majority was performed in adult study populations, and folate was the main nutrient of interest. Several candidate gene and epigenome-wide association studies reported differential DNA methylation of CpG sites in response to folate (e.g. IGF2, H19, HOX), fatty acids (e.g. PPRAGC1A, TNFα), and vitamin D (CYP24A1). Some of these observed associations were specific to life course stage (e.g. IGF2 in early life) and tissue (e.g. opposite directions for PPRAGC1A in muscle versus fat tissue).

Conclusions: To date, promising results have been reported in the field of nutrition and DNA methylation in humans at different stages across the life-course; especially for nutrients known to be involved in one-carbon metabolism, including folate, but also others, such as fatty acids and vitamin D. Studies on other nutrients, such as other macronutrients and several minerals are still scare. Further large-scale studies of high quality are needed to expand our understanding on the role of nutrition in DNA methylation and its effects on health and disease.

1. INTRODUCTION

DNA methylation is prone to modification by environmental factors, including nutrition. The susceptibility of change in DNA methylation in response to nutrition is particularly high during early life (1). Nevertheless, nutrition has also been reported to be associated with DNA methylation in other stages of the life course, for instance, during adolescence and adulthood (2).

Nutrients that are known to be involved in DNA methylation through their role in one-carbon metabolism are B-vitamins and methionine. However, also other nutrients such as fatty acids, protein, and vitamin D, are suggested to have an effect on DNA methylation (3). To date, many studies have been carried out investigating the role of nutrition on DNA methylation, in animal studies as well as human studies. Although evidence from animal studies demonstrates that nutrition has effects on DNA methylation, findings from studies in humans are inconsistent (4).

In order to gain more insight in effects of several nutrients on DNA methylation across the life-course, a clear overview of the current knowledge is of importance. Identifying which nutrients affect DNA methylation, either globally or at specific CpG sites, will provide insight in the mechanisms that are responsible for the effect of nutrition on several health outcomes. Therefore, we conducted a comprehensive systematic review on the relationship between status and intake of nutrients with DNA methylation in humans across the life course.

2. METHODS

This systematic review was performed and reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (5).

Literature search

A literature search was designed for six electronic databases by an experienced biomedical information specialist. The search engines Embase.com (Medline and Embase), Medline (Ovid), Cochrane Central, Web-of-Science, PubMed, and Google Scholar were searched from inception until May 10th 2016 (date last searched) to identify published studies that examined the association between nutrients and DNA methylation. The full search strategies of all databases are provided in **Supplement 1**.

Study selection and inclusion criteria

We selected studies that examined the association between nutrients (blood levels; dietary intake; and/or dietary supplements) and DNA methylation (global, site specific, and/or genome-wide) in humans. We excluded studies that were performed among only patients with chronic diseases (e.g. Alzheimer's disease, diabetes, anorexia nervosa, cardiovascular diseases, etc.) and case reports (n<5). No restrictions were set on year of publication, language, or study design. We excluded studies on caloric intake, alcohol intake, glucose levels, triglyceride levels, and cholesterol levels as these

were outside the scope of our review. Two reviewers screened the retrieved titles and abstracts and selected eligible studies according to predefined selection criteria independently of each other (**Supplement 2**). Discrepancies between the two reviewers were resolved through discussion, with an arbitrator available if no consensus was reached. We retrieved full texts for studies that satisfied all selection criteria. These full-text articles were evaluated in detail once more by two investigators against the selection criteria.

Data extraction

A structured database was developed prior to the data extraction. Detailed characteristics of individual studies were extracted including study design, study size, country, characteristics of the study population, and details on exposure and outcome assessment. In addition, we extracted information on covariate adjustments, and conclusions. The association of each nutrient with each methylation measure, either global or genome-wide site-specific or gene-specific, was extracted separately to report each specific analysis. Gene-specific studies were considered per gene separately.

Quality analysis

The quality of included studies was evaluated by two reviewers using a predefined scoring system. This quality score (QS) was previously developed for its use in systematic reviews and meta-analyses including studies with various study designs (6). A score of 0, 1 or 2 points was allocated to each of the following five items: 1) study design; 2) size of the population for analysis; 3) quality of the methods used for exposure assessment or appropriate blinding of an intervention; 4) quality of the methods used for outcome assessment; and 5) adjustment for potential confounders or adequate randomization of an intervention. The combined scores resulted in a total QS between 0 and 10 points, with 10 representing the highest quality. Details on the QS are presented in Supplement 3.

3. RESULTS

Characteristics of the included studies

From the literature search we identified 3,774 unique references, of which 3,523 were excluded after screening of title and abstract based on the selection criteria. Of the 251 remaining references, full-texts were retrieved and reviewed, of which 98 studies met all criteria and were included in this systematic review (**Figure 1**). Of the included studies, 25 studies investigated the association between maternal nutrition and offspring DNA methylation, nine studies were carried out during infancy, childhood, or adolescence, and 70 studies were performed in adults (**Table 1**, **Figure 2**). Most studies examined associations between nutrients and gene-specific DNA methylation, of which the majority focused on folate as nutrient of interest (**Tables 2-18**). Summaries of the

findings of all included studies are presented below per nutrient or nutrient groups and per life stage. Detailed information of population characteristics, DNA methylation assessment, confounder adjustment, and description of results is included in the online supplementary material.

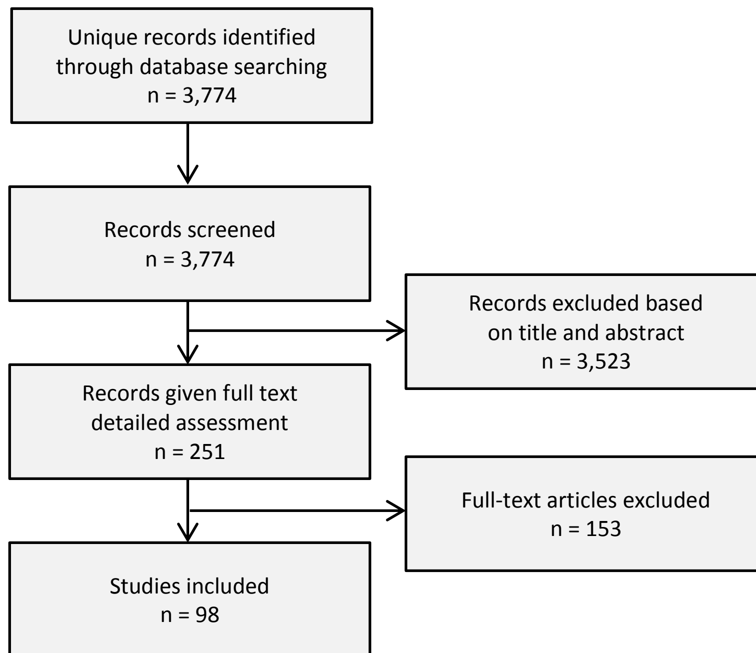


Figure 1. Flow chart of included studies

Table 1. Number of studies per category of life course stage, nutrient group, and DNA methylation type

	Maternal nutrition & offspring DNA methylation				**Infant, child, & adolescent nutrition & DNA methylation**				**Adult nutrition & DNA methylation**				
	Total	**Global**	**Gene-specific**	**Genome-wide**	**Total**	**Global**	**Gene-specific**	**Genome-wide**	**Total**	**Global**	**Gene-specific**	**Genome-wide**	**TOTAL**
MICRONUTRIENTS													
Vitamin A, α-carotene & β-carotene	0	0	0	0	1	1	0	0	3	1	2	0	4
Folate	16	6	7	4	3	3	1	0	43	34	12	1	59
Vitamin B1	0	0	0	0	0	0	0	0	1	1	0	0	1
Vitamin B2	1	0	1	0	0	0	0	0	5	5	0	0	6
Vitamin B3	1	0	1	0	0	0	0	0	1	1	0	0	2
Vitamin B6	1	0	1	0	0	0	0	0	11	11	1	0	12
Vitamin B12	4	2	2	0	2	1	1	0	22	18	7	0	26
Vitamin C	0	0	0	0	0	0	0	0	5	1	4	0	5
Vitamin D	3	0	2	1	3	1	2	1	5	3	3	0	10
Vitamin E	0	0	0	0	0	0	0	0	4	1	3	0	4
Choline & betaine	2	2	1	0	0	0	0	0	3	3	1	0	3
Minerals & trace elements	0	0	0	0	1	1	0	0	7	6	2	0	8
Combined nutrients	1	0	0	1	0	0	0	0	6	3	2	1	7
Bioactive compounds	0	0	0	0	0	0	0	0	6	1	5	0	6
MACRONUTRIENTS													
Fat & fatty acids	5	2	3	1	2	0	0	2	18	4	10	5	25
Carbohydrates & fiber	2	1	1	0	0	0	0	0	6	4	3	0	8
Protein & amino acids	2	1	1	0	0	0	0	0	7	5	3	0	9
TOTAL	25	9	13	6	9	4	4	3	70	42	26	7	**98**

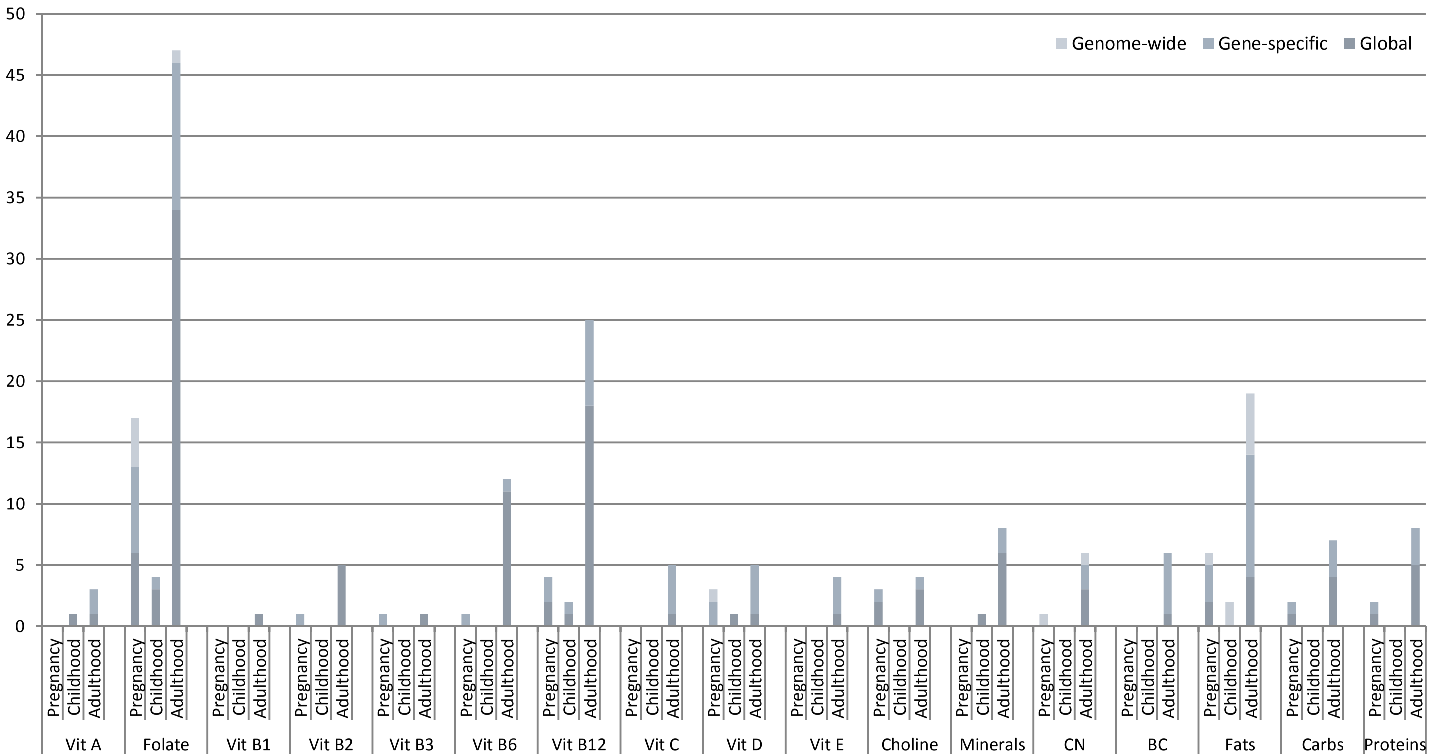


Figure 2. Distribution of studies (N=98) per category of life course stage, nutrient group, and DNA methylation outcome type. Vitamin (Vit) A, includes carotenoids; Choline, includes choline and betaine; Minerals, include minerals and trace elements; CN, Combined nutrients; BC, Bioactive compounds; Fats, include fats and fatty acids; Carbohydrates (Carbs), include fibers; Proteins, include proteins and amino acids.

MICRONUTRIENTS

Vitamin A, α-carotene, and β-carotene

Vitamin A in infancy, childhood, and adolescence and DNA methylation

Perng et al (QS: 6) showed that higher plasma vitamin A levels in children aged 5 to 12 years were associated with global DNA hypomethylation in blood (7).

Vitamin A in adulthood and DNA methylation

Piyathilake et al. (QS: 6) observed no association between plasma vitamin A levels and global DNA methylation in either PBMCs or cervical cells (8). Bollati et al (QS: 6) investigated the association of intake of α-carotene, β-carotene, and retinol with methylation of CD14, Et-1, HERV-w, iNOS and TNFα in blood. They observed an association of higher intake of β-carotene with hypermethylation of HERV-w and higher intake of β-carotene and retinol was associated with hypomethylation of TNFα. However, no significant association was observed for the other genes (9). Stidley et al (QS: 6) did not find an association of α-carotene, β-carotene, and retinol intake with methylation index of genes (p16, MGMT, DAPK, RASSF1A, PAX5α, PAX5β, GATA4, and GATA5) in sputum (10).

Folate

Maternal folate and offspring DNA methylation

Sixteen unique studies investigated the association between maternal folate and offspring DNA methylation, six at a global level, seven gene-specific, and four at a genome-wide level. Of the four studies investigating global DNA methylation in cord blood (QS: 2-6), one found that an increased dietary folate as well as RBC folate were associated with global hypomethylation (11), whereas the other three studies reported no significant association with folate (12-14). In contrast, studies investigating DNA methylation in other fetal tissues, including placenta, brain and heart, showed that higher maternal folate levels were associated with global hypermethylation (15, 16). Four studies investigated the association between levels or intake of maternal folate and IGF2 methylation in offspring (QS: 5-7). Three of these studies measured methylation in cord blood. In these studies, higher folate intake was shown to be associated with IGF2 hypermethylation (11) and no association was observed between serum folate and DNA methylation (17). Higher RBC folate showed association with IGF2 hypermethylation in one study (18) and no association in another study (11). One observational study found IGF2 hypermethylation in offspring at 12-18 months in mothers who used folate supplements during pregnancy compared to those of mothers who did not (19). Furthermore, Hoyo et al. (QS: 5 and 7) found an association of maternal folate supplements and erythrocytes with H19 hypomethylation in cord

blood (18, 20). Maternal intake and erythrocyte folate were found to be associated with PEG3 hypomethylation in cord blood (11, 18), whereas no association was found with maternal RBC folate (QS: 5-7) (11). Furthermore, higher maternal erythrocyte folate was also associated with hypermethylation of PEG1/MEST and hypomethylation of PEG10/SGCE in cord blood (QS: 7) (18). Higher maternal erythrocyte folate was associated with hypomethylation of MEG3, MEG3-IG (intergenic), NNAT and PLAGL1 promoters in cord blood (QS:7) (18). Van Mil et al. showed that supplement use and venous blood levels of folate were associated with NR3C1 hypermethylation, and higher folate supplement use was associated with 5-HTT hypomethylation in cord blood (QS: 6 and 7) (21). Furthermore, no associations were observed between maternal folate intake or RBC and SNRPN or ZAC1 methylation in cord blood (QS: 5 and 6) (11, 22).

Of the four studies investigating maternal folate levels and epigenome-wide DNA methylation in cord blood, three used Illumina 450K arrays and identified several CpG sites. Amarasekera et al. (QS: 4, N=23) found seven differentially methylated regions, of which ZFP57 was validated using Sequenom EpiTyper platform (23). Joubert et al. (QS: 8, N=1988) identified 443 CpG sites annotated to 320 genes, of which some novel genes included APC2, GRM8, SLC16A12, OPCML, PRPH, LHX1, KLK4 and PRSS21 (24). Gonseth et al. (QS: 6, N=347) found that maternal folate intake was associated with three CpG sites annotated to genes TFAP2A, STX11 and CYS1 (25). The fourth EWAS (QS: 6, N=200) did not find associations between maternal folate levels and genome-wide DNA methylation (26).

Folate in infancy, childhood, and adolescence and DNA methylation

Three unique studies investigated the association between folate levels and global DNA methylation during infancy or childhood, of which one also examined gene-specific DNA methylation. Two studies (QS:4 and 5) examined the association between folate levels and global DNA methylation in cord blood: Haggerty et al. showed that RBC folate was associated with global hypomethylation whereas Fryer et al. observed no association (11, 13). Perng et al. (QS: 4) studied the association between erythrocyte folate and global methylation in children aged 5-12 years, but found no association (7). Haggerty et al. (QS: 6) also examined the association between RBC folate and DNA methylation of IGF2, PEG3 and SNRPN in cord blood. In line with their findings for maternal folate and offspring DNA methylation, they found that RBC folate at birth was also associated with IGF2 hypermethylation, PEG3 hypomethylation and no difference in SNRPN methylation (11).

Folate in adulthood and DNA methylation

There were 41 unique studies that investigated the association between folate and adult DNA methylation. Thirty-three studies examined DNA methylation at a global

level, 12 gene-specific, and one at a genome-wide level. In observational studies that measured DNA methylation in blood, folate intake was associated with DNA hypomethylation in one study (QS: 5) (27), DNA hypermethylation in three studies (QS: 3-5) (28-30), but no difference in methylation was observed in majority of the studies (QS: 1-6) (12, 31-35). Higher blood levels of folate were associated with DNA hypomethylation in one study (QS: 6) (36), whereas DNA hypermethylation in six studies (QS: 4-6) (37-42). A few studies showed no difference in methylation (QS: 3-6) (34, 43-45). Overall, dietary intake and levels of folate in these observational studies tend to be more often associated with global DNA hypermethylation than hypomethylation. In contrast, intervention studies with folate supplements showed an effect of folate on DNA hypomethylation in three studies (QS: 2-8) (46-48), DNA hypermethylation in two studies (QS: 3) (49, 50) and no effect on methylation in three studies (QS: 3-8) (51-53). The direction of effect in these intervention studies were contradictory to those shown in the observational studies. In studies that measured DNA methylation in colonic tissue, higher folate intake was associated with global DNA hypomethylation on one side of the colon (QS: 3) (54), whereas folate in blood levels (54-57) and colonic tissue (55) showed no association with DNA methylation (QS: 4-6). Furthermore, Llanos et al. measured DNA methylation in breast tissue and observed no association with folate levels from plasma and breast tissue (QS: 6) (58). In an intervention study, Aarabi et al. found a significant association between higher folate supplements and DNA hypomethylation (QS: 4) (59).

Two studies investigated the association between levels or intake of folate and methylation in IGF2 and H19. In contrast to the studies that examined the same genes in maternal-offspring (11, 18-20), Hanks et al. (QS: 7) did not find an association between folate in serum, RBC and colonic tissue with IGF2 methylation in adults (55). In addition, in an intervention study by Aarabi et al. (QS: 5), no effect was observed between folate supplements and methylation in H19 (59).

Higher folate intake was associated with promotor hypomethylation of TNFα and methylation index of genes (p16, MGMT, DAPK, RASSF1A, PAX5α, PAX5β, GATA4, and GATA5) in blood and sputum, respectively. However, no association was observed with promoter TLR2, CD14, Et-1, HERV-w, iNOS and IL-6 methylation in blood (QS: 5-6) (9, 30, 60). Dhillon et al. (QS: 5) showed no significant association between folate intake and GSTM1 methylation in serum. However, when analysis was stratified for MTHFR genotype, low folate intake was associated with lower methylation of GSTM1 methylation for CT and TT group only (61). Two studies investigated the association of levels of folate in serum, RBC, plasma, and colonic tissue with gene-specific methylation in colonic tissue or rectal mucosa (QS: 6-7). Hypermethylation of MYOD, SFRP1, SFRP2 and methylation index of genes (HPP1, APC, SFRP1, SFRP2, SOX17, WIF1, ESR1, MYOD, N33) was observed with higher folate levels. However, no significant associations were observed for methylation of individual genes: APC, MYOD1, MLH1,

N33, SOX17 and/or ESR1 (55, 57). Ottini et al. (QS: 5) reported that higher folate levels in plasma were associated with hypomethylation of methylation index of genes (p16, FHIT, RAR, CDH1, DAPK1, hTERT, RASSF1A, MGMT, BRCA1 and PALB2) (62).

Some studies observed associations for nutrient levels in a tissue specific manner. Tapp et al. (QS: 6) investigated the association of folate levels with HPP1 and WIF1 methylation in rectal mucosa. They observed hypermethylation of HPP1 and WIF1 in association with higher plasma folate levels, but not with higher RBC folate levels (57). Hanks et al. (QS: 7) examined the association between folate levels and MGMT methylation in colonic tissue. They observed hypomethylation of MGMT in association with higher serum folate levels, but not with higher folate levels in RBC and colonic tissue (55). In women with HPV, higher plasma folate levels were associated with HPV 16 hypermethylation (promoter) in blood (QS: 5) (63). In healthy women, higher plasma and breast folate levels were associated with hypomethylation of p16INK4a (QS: 7) (64). In intervention studies, folate supplementation had no effect on ESR1 and MLH1 methylation in colonic mucosa, and DLK1/GTL2, MEST, SNRPN, PLAGL1 and KCNQ1OT1 methylation in sperm (QS: 5 and 7) (59, 65). Song et al. (QS: 4) performed an EWAS of folate levels in breast tissue in women, and found two differentially methylated CpG sites. One CpG site near JAG2 was hypomethylated and another CpG site near DNAJC2 was hypermethylated (66).

Vitamin B1

Vitamin B1 in adulthood and DNA methylation

Marques-Rocha et al. (QS: 5) showed that higher vitamin B1 intake was associated with global DNA hypomethylation in blood (67).

Vitamin B2

Maternal vitamin B2 and offspring DNA methylation

Azzi et al. (QS: 6) showed that higher vitamin B2 intake during pregnancy was associated with hypermethylation of ZAC1 in cord blood (22).

Vitamin B2 in adulthood and DNA methylation

Five studies (QS: 3-5) investigated the association between vitamin B2 intake and global DNA methylation in blood or colonic tissue (27, 30, 35, 54, 67), and only one study found an association: that higher vitamin B2 intake was associated with DNA hypermethylation in blood (67). In addition, Figueiredo et al. (QS: 4) found no association between vitamin B2 levels in plasma with global methylation in colonic tissue (54). Zhang et al. (QS: 5) showed that vitamin B2 intake was not associated with methylation of IL-6 in blood (30).

Vitamin B3

Maternal vitamin B3 and offspring DNA methylation

Azzi et al. (QS: 6) showed that vitamin B3 intake during pregnancy was not associated with methylation of ZAC1 in cord blood (22).

Vitamin B3 in adulthood and DNA methylation

Marques-Rocha et al. (QS: 5) showed that higher vitamin B3 intake was associated with global DNA hypomethylation in blood (67).

Vitamin B6

Maternal vitamin B6 and offspring DNA methylation

Azzi et al. (QS: 6) reported that vitamin B6 intake during pregnancy was not associated with methylation of ZAC1 in cord blood (22).

Vitamin B6 in adulthood and DNA methylation

Eight studies (QS: 3-5) investigated the association between vitamin B6 intake with global DNA methylation in blood or colonic tissue (7, 27, 29-31, 34, 35, 54), and only one study found an association and reported that higher vitamin B6 intake was associated with global hypomethylation in blood (31). Two other studies (QS: 4 and 6) investigated the association of vitamin B6 levels in plasma and venous blood with global DNA methylation in blood or colonic tissue, but did not find any association (38, 54). An intervention study by Hübner et al. (QS: 3), showed that vitamin B6 supplementation had no effect on global DNA methylation in blood (51). Zhang et al. (QS: 5) showed that vitamin B6 intake was not associated with methylation of IL-6 in blood (30).

Vitamin B12

Maternal vitamin B12 and offspring DNA methylation

Two studies investigated the association of intake and serum levels of vitamin B12 with global DNA methylation in cord blood (QS: 5 and 6). Higher serum levels of vitamin B12 were associated with hypermethylation (14), whereas no difference in methylation was observed with higher vitamin B12 intake (12). For the gene-specific studies, Ba et al. (QS: 6) showed that higher vitamin B12 levels in serum were associated with IGF2 promoter hypomethylation in cord blood (17). Azzi et al. (QS: 6) showed that higher vitamin B12 intake during pregnancy was associated with hypermethylation of ZAC1 in cord blood (22).

Vitamin B12 in infancy, childhood, and adolescence and DNA methylation

In children aged 5 to 12 years, no association was observed between plasma vitamin B12 levels and global DNA methylation in blood (QS: 4) (7). In cord blood, McKay et al. (QS: 7) showed that higher serum vitamin B12 levels were associated with hypomethylation of IGFBP3 (14).

Vitamin B12 in adulthood and DNA methylation

Eight studies (QS: 3-6) found no association between vitamin B12 intake and global DNA methylation (12, 27, 29-32, 34, 35). Also for vitamin B12 levels in plasma or serum, no associations were found with global DNA methylation in blood, colonic tissue or rectal mucosa in eight studies (QS: 3-6) (34, 38, 43-45, 54, 56, 57). However, in one study (QS: 6), higher plasma vitamin B12 levels were associated with global hypermethylation in blood clots (68). An intervention study (QS: 3) showed that vitamin B12 supplements had no effect on global DNA methylation in blood (69). For gene-specific studies, vitamin B12 intake was not associated with methylation of CD14, Et-1, HERV-w, iNOS, TNFα, IL-6, and TLR2 in blood (QS: 5-6) (9, 30, 60). In addition, Tapp et al (QS: 6) found no association between vitamin B12 levels in plasma and methylation of HPP1, APC, SFRP1, SFRP2, SOX17, WIF1, ESR1, MYOD, N33 and methylation index of these genes (57). Al-Ghnaniem et al. (QS: 6) observed that higher vitamin B12 levels in venous blood were associated with hypomethylation of ERα in colonic mucosa (70). Furthermore, plasma vitamin B12 levels were associated with hypermethylation of HPV 16 (QS: 5) and no difference in methylation index of genes (p16, FHIT, RAR, CDH1, DAPK1, hTERT, RASSF1A, MGMT, BRCA1 and PALB2) (QS: 3) (62, 63).

Vitamin C

Vitamin C in adulthood and DNA methylation

Piyathilake et al. (QS: 6) investigated the association between plasma vitamin C levels and global DNA methylation in PBMCs and cervical cells, but did not find a significant association (8). Two studies (QS: 6) investigated the association between vitamin C intake and gene-specific promoter DNA methylation. One study found that higher intake of vitamin C was associated with hypomethylation of PON1 in venous blood (71), whereas the other study did not find an association of vitamin C intake with methylation index of genes (p16, MGMT, DAPK, RASSF1A, PAX5α, PAX5β, GATA4, and GATA5) in sputum (10). Furthermore, Bollati et al (QS: 6) investigated the association of intake of ascorbic acid with methylation of CD14, Et-1, HERV-w, iNOS and TNFα in blood, but no significant association was observed (9). Piyathilake et al. (QS: 5) found no association between vitamin C levels in plasma and HPV 16 methylation in blood (63).

Vitamin D

Vitamin D and offspring DNA methylation

One study (QS: 5) found that higher maternal vitamin D levels in plasma were associated with hypomethylation of RXRA in umbilical cord tissue (72), whereas another study (QS: 4) did not find an association between maternal vitamin D levels in serum and CYP24A1 methylation in placenta (73). Mozhui et al (QS: 6) conducted an EWAS of vitamin D levels, but found no significant CpG site associations in cord blood (26).

Vitamin D in infancy, childhood, and adolescence and DNA methylation

For vitamin D, Zhu et al. (QS: 7) found that higher plasma vitamin D levels as well as an intervention with vitamin D supplementation were associated with global hypermethylation in blood (74). In neonates, Novakovic et al (QS: 4) found no association between serum vitamin D levels and CYP24A1 methylation in placenta (73). Zhu et al. (QS: 4) investigated the associated between plasma vitamin D levels and gene-specific methylation in blood among children around 16 years of age. Higher vitamin D levels were found to be associated with hypomethylation of CYP2R1, hypermethylation of CYP24A1, and differential methylation in opposite directions of two CpG sites annotated to DHCR7 (75). In an EWAS of vitamin D levels, Zhu et al. (QS: 5) identified two differentially methylated CpG sites: higher vitamin D levels were associated with hypomethylation of DOI3 and hypermethylation of MAPRE2 (75).

Vitamin D in adulthood and DNA methylation

Three studies, of which two were cross-sectional (QS: 5) and one intervention (QS: 3) found no effect of vitamin D supplementation or vitamin D levels in plasma or serum on global DNA methylation in blood or rectal mucosa (51, 57, 76). In gene-specific studies, Bollati et al (QS: 6) found no association between vitamin D intake and promoter methylation of CD14, Et-1, HERV-w, iNOS and TNFα (9). Tapp et al. (QS: 6) investigated the association between plasma vitamin D levels and gene-specific methylation of HPP1, APC, SFRP1, SFRP2, SOX17, WIF1, ESR1, MYOD and N33 in rectal mucosa. They found that higher vitamin D levels were associated with hypomethylation of index of these genes and also individual genes such as APC, WIF1 and MYOD. No association was found for the other individual genes (57). Furthermore, Ashktorab et al. (QS: 7) found no significant association between serum vitamin D levels and DKK1 methylation in blood (77).

Vitamin E

Vitamin E in adulthood and DNA methylation

Piyathilake et al. (QS: 6) observed that higher levels of plasma vitamin E were associated with global hypomethylation in PBMCs, but not with global methylation in cervical cells (8). In a gene-specific study, Stidley et al (QS: 6) did not find an association of vitamin E intake with methylation index of any of the studied genes (p16, MGMT, DAPK, RASSF1A, PAX5α, PAX5β, GATA4, and GATA5) in sputum (10).

Choline and Betaine

Maternal choline and betaine and offspring DNA methylation

Boeke et al. (QS: 6) showed that higher choline or betaine intake during pregnancy were associated with global DNA hypomethylation in cord blood (12). In contrast, Jiang et al. (QS: 6), in an intervention study, reported that higher supplemental choline intake during pregnancy was associated with global DNA hypermethylation in placenta, but not in cord blood (78). For gene-specific studies, an intervention study by Jiang et al. (QS: 5) showed that higher supplemental choline intake during pregnancy was associated with hypermethylation of promoter CRH and NR3C1 in placenta. In contrast to placental tissue, in cord blood promoter CRH and NR3C1 were hypomethylation in response to a higher supplemental choline intake. No effect of choline supplementation was observed in methylation of GNAS-AS1, IGF2, IL-10, or LEP in placenta or cord blood (78).

Choline and betaine in adulthood and DNA methylation

Shin et al. (QS: 3) showed that choline supplementation was associated with global DNA hypermethylation in blood, but only for subjects with the MTHFR 677CC genotype (79). However, another intervention study (QS: 4) showed no effect of choline supplementation on global DNA methylation in blood (78). In addition, in a cross-sectional study by Boeke et al. (QS: 6), choline and betaine intake was not associated with global DNA methylation in blood in pregnant women (12). In an intervention study by Jiang et al. (QS: 5) a higher supplemental choline intake was not associated with methylation of promoter CRH and NR3C1 in blood among pregnant women (78).

Minerals and trace elements

Minerals and trace elements in infancy, childhood, and adolescence and DNA methylation

In a cross-sectional study among children aged 5 to 12 years, Perng et al (QS: 4) investigated the association of plasma levels of ferritin and serum levels of zinc with global DNA methylation in blood, but no significant associations were observed (7).

Minerals and trace elements in adulthood and DNA methylation

Marques-Rocha et al. (QS: 5) showed that higher intakes of both copper and iron were associated with global DNA hypermethylation in blood (67). In contrast, Gomes et al (QS: 3) reported that higher intakes of magnesium were associated with DNA hypomethylation in blood (31). Two studies (QS: 3 and 5) examined associations between zinc intake and global DNA methylation in blood, but both observed no significant associations (31, 32). Furthermore, Tapp et al. (QS: 5) reported that higher plasma levels of selenium were associated with global DNA hypermethylation in rectal mucosa, but only in women (57). McChelland et al. (QS: 6) reported that higher serum levels of phosphate were associated with global DNA hypomethylation in blood (80). In an intervention study by Hubner et al. (QS: 3) no effect was observed of calcium supplementation on global DNA methylation in blood (51).

For gene-specific studies, Shimazu et al. (QS: 6) investigated the associations of salt intake with methylation of miR-124a-3, EMX1 and NKX6-1 in gastric mucosa, but observed no significant associations (81). Tapp et al. (QS: 6) examined associations of plasma selenium levels with methylation of HPP1, APC, SFRP1, SFRP2, SOX17, WIF1, ESR1, MYOD, N33 and an index of these genes. They observed that higher levels of selenium were associated with hypermethylation of N33, but only in males. No significant associations were observed for the other genes or index (57).

Combined nutrients

Maternal combined vitamins and minerals and offspring DNA methylation

In an intervention study, Khulan et al. examined the effect of maternal micronutrient supplementation on genome-wide DNA methylation in cord blood (QS: 6) and peripheral blood during infancy (QS:6) (82). In cord blood 14 CpG sites were identified in males and 21 CpG sites in females, with no overlap between the two. In addition, in infant blood 108 CpG sites were identified in males and 106 CpG sites in females, with only 5 (XIRP1, MEOX1, GNA11, C1orf54, KRTAP21-1) in agreement between male and female infants. Within each sex a significant proportion of the identified CpG sites in cord bloods were also observed in infants (males 7/14: MEOX1, SPAG4L, SPATA22, NRN1L, C1orf54, PIP, PTPN20B; females 8/21: HSPC176, XIRP1, BPIL1, MGMT, CHIT1, C14orf152, KLRC2, DEFB123).

Combined vitamins and minerals in adulthood and DNA methylation

Two studies investigated the association between folate and vitamin B12 combined with global DNA methylation. Piyathilake et al. (QS: 6) reported that higher vitamin B12 and folate in plasma were associated with global hypermethylation in PBMCs, but not in cervical cells (8). However, in an intervention study by Fenech et al. (QS: 6) vitamin B12 and folate supplements had no effect on global DNA methylation (43). Pusceddu et al. (QS:5) studied the effect of vitamin B12, vitamin B6, folate, vitamin D and calcium

supplementation on global DNA methylation After intervention LINE-1 hypermethylation was observed (83). For gene-specific studies, Stidley et al. (QS: 6) observed that higher intake of multivitamins was associated with hypomethylation of methylation index (p16, MGMT, DAPK, RASSF1A, PAX5α, PAX5β, GATA4, and GATA5) (10). In an intervention study by Zhou et al. (QS: 5), vitamin D and calcium supplementation led to hypomethylation of CYP24A1, but had no effect on CYP2R1, CYP27A1, and CYP27B1 (84). Kok et al. studied the effect of vitamin B12 and folate supplementations on epigenome-wide DNA methylation (QS: 8). After intervention 162 CpG sites were identified as compared to baseline. Comparisons of folic acid and vitamin B12 versus placebo revealed one CpG sites (cg19380919) and 6 differentially methylated regions (DMRs), with pronounced changes in DIRAS3, ARMC8, and NODAL. In addition, serum levels of folate and vitamin B12 were related to DNA methylation of 173 and 425 regions, respectively, including several members of the developmental HOX genes (85).

Bioactive compounds

Bioactive compounds in adulthood and DNA methylation

Bollati et al (QS: 6) investigated the association of intake carotenoids, polyphenols, and flavonoids with methylation of CD14, Et-1, HERV-w, iNOS and TNFα in blood. They observed an association of higher intake of carotenoids with hypermethylation of HERV-w and with hypomethylation of TNFα. However, no significant association was observed for the other genes (9). De la Iglesia et al. (QS: 6) observed that that higher lycopene intake was associated with PON1 hypomethylation (71). Zhong et al. (QS: 6) reported that flavonoid intake was associated with TLR2 hypomethylation (60). Stidley et al (QS: 6) did not find an association of carotene, lutein, zeaxanthin, and lycopene intake with methylation index of genes (p16, MGMT, DAPK, RASSF1A, PAX5α, PAX5β, GATA4, and GATA5) in sputum (10).

MACRONUTRIENTS

Fat and fatty acids

Maternal fat and fatty acids and offspring DNA methylation

Rerkasem et al. (QS: 3) found no association between total fat intake during pregnancy and global DNA methylation in blood of children around 20 years of age (86). Godfrey et al. (QS: 3) also found no association between total fat intake during pregnancy and methylation of RXRA and eNOS in cord blood (87). For individual fatty acids, an intervention study by Lee et al. (QS: 8) found that higher ω-3 PUFA supplementation during pregnancy had an effect on global hypermethylation in cord blood (88). In the same study, Lee et al. (QS: 8) investigated gene-specific DNA methylation and found

that higher ω-3 PUFA supplementation during pregnancy had an effect on hypermethylation of IGF2, but no association was found for methylation of IFNγ, TNFα, IL-13, GATA3, STAT3, IL-10 and FOXP3 in cord blood (88). In an intervention by Amarasekera et al. (QS: 8), no significant effects of ω-3 PUFA were found on genome-wide DNA methylation in cord blood (89).

Maternal fat and fatty acids in infancy, childhood, and adolescence and DNA methylation

Voisin et al. (QS: 5) conducted an EWAS and identified four, 130, 158 and 16 CpG sites in blood significantly associated with intake of total fat, (MUFA+PUFA)/SFA, MUFA/SFA and PUFA/SFA, respectively, in children around the age of 10 years (90). On the other hand, in an intervention study by Lind et al. (QS: 8) found no epigenome-wide CpG site associations of ω-3 PUFA supplementation as fish oil (91).

Fat and fatty acids in adulthood and DNA methylation

Total fat: Four studies (QS: 3-4) investigated the association between total fat intake and global DNA methylation in blood, of which two found that higher fat intake was associated with global DNA hypomethylation (31, 34), whereas the other two studies did not find an association (30, 35). Gómez-Uriz et al. (QS: 4) found that higher total fat intake was associated with TNFα hypomethylation in blood (92). However, Bollati et al. and Zhang et al (QS: 5 and 6) investigated the association of intake of total fat with methylation of CD14, Et-1, HERV-w, iNOS, TNFα and IL-6 in blood, but no significant association was observed (9, 30). In addition, Stidley et al. (QS: 6) found no association of total fat as well as animal fat intake with methylation index of genes (p16, MGMT, DAPK, RASSF1A, PAX5α, PAX5β, GATA4, and GATA5) in sputum (10). Brøns and Gillberg et al. (QS: 6) investigated the effect of high fat overfeeding in healthy men on promoter PPARGC1A methylation (93, 94). They found that high fat overfeeding in men with normal birthweight was associated with PPARGC1A hypomethylation in muscle tissue (93), whereas high fat overfeeding in men with low birthweight was associated with PPARGC1A hypermethylation in subcutaneous adipose tissue (94). In an EWAS by Irvin et al. (QS: 8), fatty acids were not associated with epigenome-wide DNA methylation after correcting for multiple testing (69). In an intervention study among healthy men, Jacobsen and Gillberg et al (QS: 6) investigated the effect of high fat overfeeding on epigenome-wide DNA methylation. They observed that high fat overfeeding affected methylation at 652 CpG sites (including in CDK5, IGFBP5 and SLC2A4) in subcutaneous adipose tissue (95). Furthermore, change in DNA methylation of 7,909 CpG sites (including, DNM2, MGMT, SLC2A3/GLUT3, MRC1 and ACAT2) in skeletal muscle biopsies was observed in response to high fat overfeeding. Within the same study, it was shown that these changes in genome-wide DNA methylation were more

pronounced in participants with a low birth weight, compared to those with a normal birthweight (96).

Individual fatty acids: Zhang et al. (QS: 4) found no association between PUFA intake and global DNA methylation in blood (30, 35). For methylation measured in blood, one study found that higher ω-6 PUFA intake was associated with hypomethylation of TNFα (97) and another study found that higher PUFA intake was associated with hypermethylation of CLOCK (98). However, Zhang et al (QS: 5) and Bollati et al (QS: 6) observed no association of intake of total PUFA and ω-3 PUFA with methylation of CD14, Et-1, HERV-w, iNOS, TNFα and IL-6 in blood (9, 30). Furthermore, Ma et al. (QS: 7) found that higher ω-3 PUFA levels in erythrocytes were associated with hypomethylation of IL-6 promoter (99). In another study by Ma et al. (QS: 7), higher levels of erythrocyte EPA were associated with hypomethylation of ABCA1 promoter. In addition, higher erythrocyte ALA was associated with hypomethylation of APOE for the CC genotype, whereas higher erythrocyte ALA was associated with hypermethylation of APOE for the AA genotype (100). In an EWAS by Aslibekyan et al. (QS:6), higher levels of ω-3 PUFA in erythrocytes were associated with 27 CpG sites (including NAV1, CCL17, ACTA2/FAS and AHRR) in blood. Zhang et al. (QS: 4) found no association between MUFA intake and global DNA methylation in blood (30, 35). Milagro et al. (QS: 4) found that higher intake of MUFA was associated with hypomethylation of CLOCK in blood (98). However, Zhang et al. (QS: 5) and Bollati et al (QS: 6) investigated the association of intake of MUFA with methylation of CD14, Et-1, HERV-w, iNOS, TNFα and IL-6 in blood, but no significant association was observed (9, 30). Zhang et al. (QS: 4) observed that higher SFA intake was associated with global DNA hypomethylation when methylation was measured using pyrosequencing (35), whereas no association was observed when methylation was measured using PCR technique (30). Zhang et al. (QS: 5) and Bollati et al (QS: 6) investigated the association of intake of SFA with methylation of CD14, Et-1, HERV-w, iNOS, TNFα and IL-6 in blood, but no significant association was observed (9, 30).

Carbohydrates and fiber

Maternal carbohydrates and fiber and offspring DNA methylation

Rerkasem et al. (QS: 3) investigated the association between intake of carbohydrates during pregnancy and global DNA methylation in their offspring around the age of 20 years, but observed no significant association (86). Godfrey et al. (QS: 3) explored the association between carbohydrate intake during pregnancy and gene-specific promoter DNA methylation in cord tissue. They observed that higher intake of carbohydrate was associated with RXRA hypomethylation, but no significant association was observed with eNOS methylation (87).

Carbohydrates and fiber in adulthood and DNA methylation

Four studies investigated the association between carbohydrate intake and global DNA methylation in blood. One study (QS:3) reported that higher intake of carbohydrate was associated with global DNA hypomethylation (31), whereas the other three studies (QS:4) observed no significant associations (30, 34, 35). Three studies (QS: 4-6) were conducted on the association of carbohydrate and fiber intake and gene-specific DNA methylation in blood, of which one observed that higher carbohydrate intake was associated with hypermethylation of BMAL1 (101), but in the other two studies no significant associations were observed with CD14, Et-1, HERV-w, iNOS, TNFα, and IL-6 (9, 30).

Protein and amino acids

Maternal protein and amino acids and offspring DNA methylation

Rerkasem et al. (QS: 3) observed no associations between intake of protein during pregnancy and global DNA methylation in their offspring around the age of 20 years (86). Godfrey et al. (QS: 3) found no association of protein intake during pregnancy with promoter methylation of RXRA or eNOS in cord tissue (87).

Protein and amino acids in adulthood and DNA methylation

Three studies (QS: 3-4) investigated the association between protein intake and global DNA methylation in blood, but no significant associations were observed (30, 31, 35). In addition, the four studies (QS: 3-4) that explored associations between methionine intake and global DNA methylation in blood showed no significant associations (29, 30, 32, 35). Two studies (QS: 5 and 6) investigated associations of protein intake and methylation of CD14, Et-1, HERV-w, iNOS, TNFα, and IL-6 in blood, but reported no significant associations (9, 30). Similarly, two studies (QS: 5 and 6) that examined the associations of methionine intake with IL-6 and TLR2 promoter methylation in blood showed no significant associations (30, 60).

4. DISCUSSION

This is the first systematic review on nutrients and DNA methylation in humans. This review included studies examining associations between all nutrients with different types of methylation outcomes in all life-course stages. The majority of the studies performed on this topic investigated the association of folate, other B-vitamins, or fatty acids with DNA methylation, whereas studies for other nutrients are scarce or even absent, such as for vitamin K. The evidence so far shows that folate and fatty acids are associated with DNA methylation on a global, gene-specific, and genome-wide level. Several candidate gene and epigenome-wide association studies reported differential DNA methylation of CpG sites in response to folate (e.g. IGF2, H19, HOX), fatty acids (e.g. PPRAGC1A, TNFα), and vitamin D (CYP24A1). Some of these observed associations

were specific to life course stage (e.g. IGF2 in early life) and tissue (e.g. opposite directions for PPRAGC1A in muscle vs. fat tissue). However, the directions of these associations are inconsistent across the different studies. Reasons that may explain these discrepancies in findings include the differences in the type of tissue used to measure nutrient levels as well as DNA methylation and differences across life course stages.

DNA methylation across the life course

Age affects methylation patterns via transcriptional regulation (102). Evidence from our systematic review, suggests age-specific differences in folate associations with global DNA methylation in blood. In early life, higher folate intakes and levels were associated with global DNA hypomethylation (11), whereas during adulthood higher folate intakes and levels were associated with DNA hypermethylation (28-30, 37-42). This pattern could not be observed for the other nutrients due to the limited number of studies across the different life course stages or null associations, but we expect similar patterns for other nutrients. Indeed, some studies on gene-specific DNA methylation observed similar age-specific differences. Methylation of some genes in blood may be specifically prone to environmental influences during early life. For example, evidence from our systematic literature review shows that during pregnancy, at birth and during childhood, higher folate intake and levels were associated with IGF2 hypermethylation (11, 18, 19). However, when this association was studied in adults, no significant associations were observed (55). Possibly, methylation of this gene is no longer susceptible to changes during adulthood. Similarly, higher choline intake during pregnancy was associated with CRH and NR3C1 hypomethylation in cord blood, but no association was found within pregnant women (78). Several studies included in our review suggested that IGF2 methylation could also be affected by vitamin B12 and PUFAs (17, 103). Higher folate during pregnancy and at birth was associated with PEG3 hypomethylation in cord blood (11, 18). Due to the absence of studies in adults, we cannot determine age-related associations in relation to these nutrients. Future studies should be conducted in order to establish the life course patterns of these nutrients on DNA methylation.

DNA methylation across nutrients

We were interested in exploring how different nutrients may affect methylation of the same genes, because this can be indicative of which genes are susceptible to nutritional influences in general. As it is difficult to compare findings across studies due to the different population structure, lab techniques, batch effects and analytical methods used, we compared the associations of different nutrients with methylation of the same genes within the same study to discard the bias caused by heterogeneity. From these studies that explored associations between multiple individual nutrients

and gene-specific methylation for the same genes and same tissues, we observed that genes such as CD14, eNOS, ESR1, Et-1, IL-6, iNOS, N33 and SOX17 were not susceptible to any nutritional influence (9, 30, 57, 87). Methylation of some other genes, such as APC, HERV-w, HPP1, RXRA, TLR2, SFRP1, SFRP2 and methylation index of genes (p16, FHIT, RAR, CDH1, DAPK1, hTERT, RASSF1A, MGMT, BRCA1 and PALB2) showed susceptibility to only one nutrient (9, 57, 60, 87). This suggests that methylation of these genes could be responsive to one particular nutrient due to a certain pathway in which they are involved. However, for certain other genes, such as CLOCK, HPV 16, MYOD, PON1, TNFα, WIF1, ZAC1, and methylation indexes of genes (HPP1, APC, SFRP1, SFRP2, SOX17, WIF1, ESR1, MYOD and N33; p16, MGMT, DAPK, RASSF1A, PAX5α, PAX5β, GATA4 and GATA5), methylation was susceptible to multiple nutrients (9, 10, 22, 57, 63, 71, 98). Of these genes, HPV 16 and ZAC1 were hypermethylated in association with folate, vitamin B2 or vitamin B12, but not other nutrients (22, 63), suggesting a possible role in one-carbon metabolism. Overall, these findings suggest that methylation of some genes is susceptible to nutritional influences in general, whereas methylation of other genes is only responsive to specific nutrients.

DNA methylation across nutrient tissues

The tissue used to measure nutrient exposure might contribute to the nutrient-methylation association. For certain types of tissues it may be easier to detect associations of nutrients with DNA methylation, due to a better reflection of long-term status of these nutrients (104). Within studies that investigated the association between folate from different tissues or intake and same gene-specific methylation using same methylation tissue, we observed that some genes (H19 and p16INK4a (18, 20, 64), MYOD, NR3C1_a, SFRP1, SFRP2 and methylation index (HPP1, APC, SFRP1, SFRP2, SOX17, WIF1, ESR1, MYOD, N33)) showed differential methylation in association with folate regardless of the source: for plasma, RBC and dietary folate (21, 57). However, methylation of HPP1, MGMT, PEG3, and WIF1 consistently showed no association with RBC folate, but showed differential methylation in association with plasma, serum or dietary folate within the same studies (11, 55, 57). Even though RBC folate has shown associations that are in line with other folate tissues in most studies, these four studies have shown discrepancies in their findings, which might be explained by the difference in reflection of long-term folate status between RBC and other tissue of folate (105).

DNA methylation across methylation tissues

In addition to levels from different tissues or methods used to measure intake of nutrients, tissue-specificity in DNA methylation is an important factor to take into account when interpreting findings of the studies included in this review. A few studies explored associations of specific nutrients with gene-specific DNA methylation

measured in different tissues within the same subjects. For example, an intervention study by Jiang et al. showed opposite directions of association between choline intake during pregnancy and CRH or NR3C1 methylation in placenta versus cord blood, which shows that associations may differ between tissues (78). Another intervention study showed tissue-specific PPARGC1A methylation in muscle biopsy and subcutaneous adipose tissue in opposite directions in association with high-fat overfeeding diet (93, 94).

DNA methylation of genes that were replicable

Replication of results across different study populations is of utmost importance in order to confirm findings. In this systematic review, we identified different studies examining same nutrient and gene-specific methylation associations within the same stage of the life course. Different measures of folate during pregnancy were associated with hypomethylation of PEG3 and hypermethylation of IGF2 in offspring, and these findings were replicated across different studies (11, 18, 19). Methylation of TNF-α was not replicable across studies, possibly due to differences in population characteristics. One study explored this association in normal-weight young women, whereas the other study explored this association in overweight older adults, which may explain the non-replicability (9, 97).

Genome-wide DNA methylation

In this systematic review, we identified five EWASs of folate, of which four studies were conducted in relation to maternal nutrition and offspring DNA methylation (23-26), and one study in relation to nutrition in adulthood and DNA methylation (66). We investigated whether any genes whose methylation showed significant associations in gene-specific studies in relation to folate, were also found to be genome-wide significant in these EWASs. In line with the gene-specific study of Hoyo et al., who found an association of maternal folate supplements and erythrocytes with H19 hypomethylation in cord blood (18, 20), Joubert et al. in their EWAS of maternal plasma folate also found genome-wide significant hypomethylation of H19 (24). Furthermore, this EWAS found hypomethylation of SNRPN which showed null association in relation to folate intake in a gene-specific study (11). Two EWASs of vitamin D levels were conducted by Mozhui et al. (26) and Zhu et al. (75) but the former found null association and the later showed no comparison in genes with previous gene-specific studies.

Methodological considerations

In this systematic review, a comprehensive overview is provided on the current evidence on associations between nutrients and DNA methylation in humans of all ages and during all life stages. We included studies examining nutrient intake from diet and supplements, as well as nutrient levels measured in blood or other tissues.

Furthermore, we included studies using different tissues to measure DNA methylation, which enabled us to provide more insight on tissue-specific DNA methylation. The literature was designed by an experienced biomedical information specialist and multiple search engines were used, limiting the probability of undetected references. Furthermore, the selection process was performed by two independent reviewers. In order to objectively measure study quality, we used a quality score based on a predefined scoring system developed for systematic reviews. Study quality was determined by five study characteristics: study design, sample size, quality of exposure assessment, quality of outcome assessment, and confounder adjustment.

Overall, most studies are of medium to high quality. A majority of studies (28%) have a quality score 6 and 77% of the studies were of the quality score 5 or above (Figure 4, Table 1). However, many studies did not properly adjust for confounding. Even though age, sex, and sometimes lifestyle covariates were often taken into account, other factors important in DNA methylation studies, such as leukocyte proportion, were rarely included in the models. Lack of adjustment for these covariates may have led to spurious findings. As the majority of the included studies were of observational nature, causal inference from these studies is not possible. Only a few studies tested for effect modification by sex of genetic variation. This should be considered in future studies, as some studies show that associations may be different for males and females or different genotypes, such as MTHFR. There were a few EWASs included in our review, enabling the possibility to identify novel CpG sites in relation to certain nutrients. Nevertheless, with this type of approach caution must be taken when interpreting the results. As EWAS follows a hypothesis-free approach, significant findings could be the result of false positives caused by multiple testing. To solve this issue, several statistical methods could be used, such as Bonferroni correction or false detection rate (FDR). Furthermore, in these type of studies that use a hypothesis-free approach, there is a need for replication. Therefore, to ensure that findings are true-positives, a discovery and a replication cohort should be used. Even though almost all EWAS included in this review used a proper method for multiple testing, not all studies included a replication cohort in their study. Therefore, findings from these studies still need to be replicated in other cohorts in order to confirm that novel CpG sites identified in EWAS are true positives. On the other hand, the EWAS approach is also prone to type II error, or false-negatives. Usually, this is caused by lack of power due to a small sample size. Out of the included studies that used EWAS, only five had a sample size of >100 participants. Possibly, potential associations between nutrients and genome-wide DNA methylation may have been overlooked in studies with a small sample size.

CONCLUSIONS

To date, promising results have been reported in the field of nutrition and DNA methylation in humans at different stages across the life-course. In particular, nutrients known to be involved in one-carbon metabolism, such as folate, have been shown to be related to DNA methylation at a global, gene-specific and genome-wide level. In addition, also other nutrients, such as fatty acids, are suggested to be involved in DNA methylation. However, further large-scale studies of high quality are needed to expand our understanding on the role of nutrition in DNA methylation and its effects on health and disease.

Table 2. Study characteristics and results: Vitamin A, α-carotene, and β-carotene

Nutritional exposure	First author, year	Study design	N	Country	Exposure assessment	Genes	Tissue	Direction results	QS
CHILDHOOD									
GLOBAL									
Vitamin A	Perng, 2012	Cross-sectional	568	Colombia	Plasma	NA	Blood	Hypo	6
ADULTHOOD									
GLOBAL									
Vitamin A	Piyathilake, 2011	Cross-sectional	376	USA	Plasma	NA	PBMC	No	6
Vitamin A	Piyathilake, 2011	Cross-sectional	376	USA	Plasma	NA	Cervical cells	No	6
GENE-SPECIFIC									
Vitamin A (Retinol)	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*CD14*	Blood	No	6
Vitamin A (Retinol)	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*Et-1*	Blood	No	6
Vitamin A (Retinol)	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*HERV-w*	Blood	No	6
Vitamin A (Retinol)	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*iNOS*	Blood	No	6
Vitamin A (Retinol)	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*TNFα*	Blood	Hypo	6
Vitamin A (Retinol)	Stidley, 2010	Cross-sectional	1101	USA	FFQ	Index*	Sputum	No	6
Carotenoids (α-carotene)	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*CD14*	Blood	No	6
Carotenoids (α-carotene)	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*Et-1*	Blood	No	6
Carotenoids (α-carotene)	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*HERV-w*	Blood	No	6
Carotenoids (α-carotene)	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*iNOS*	Blood	No	6
Carotenoids (α-carotene)	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*TNFα*	Blood	No	6
Carotenoids (α-carotene)	Stidley, 2010	Cross-sectional	1101	USA	FFQ	Index*	Sputum	No	6
Carotenoids (β-carotene)	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*CD14*	Blood	No	6
Carotenoids (β-carotene)	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*Et-1*	Blood	No	6
Carotenoids (β-carotene)	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*HERV-w*	Blood	Hyper	6
Carotenoids (β-carotene)	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*iNOS*	Blood	No	6
Carotenoids (β-carotene)	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*TNFα*	Blood	Hypo	6
Carotenoids (β-carotene)	Stidley, 2010	Cross-sectional	1101	USA	FFQ	Index*	Sputum	No	6

*Index (*p16, MGMT, DAPK, RASSF1A, PAX5α, PAX5β, GATA4,* and *GATA5*)

Table 3. Study characteristics and results: Folate

First author, year	Study design	N	Country	Exposure assessment	Genes	Methylation Tissue	Direction results	QS
MATERNAL OFFSPRING RELATIONSHIP								
GLOBAL								
Fryer, 2009	Cross-sectional	24	UK	Questionnaire (supplements)	NA	Cord blood	No	2
Boeke, 2012	Cross-sectional	534	USA	FFQ	NA	Cord blood	No	6
Haggarty, 2013	Cross-sectional	913	UK	FFQ	NA	Cord blood	Hypo	4
Haggarty, 2013	Cross-sectional	913	UK	RBC	NA	Cord blood	Hypo	5
McKay, 2012	Cross-sectional	197	UK	RBC	NA	Cord blood	No	5
Chang, 2011	Cross-sectional	20	China	Blood	NA	Tissues[T]	Hyper	2
Park, 2005	Cross-sectional	107	Korea	Serum	NA	Placenta	Hyper	6
GENE-SPECIFIC								
Hoyo, 2011	Cross-sectional	438	USA	Questionnaire	*H19*	Cord blood	Hypo	5
Azzi, 2014	Cross-sectional	2002	France	FFQ & 24h recall	*ZAC1*	Cord blood	No	6
Azzi, 2014	Cross-sectional	2002	France	FFQ & 24h recall (supplements)	*ZAC1*	Cord blood	No	6
van Mil, 2014	Cross-sectional	463	Netherlands	FFQ (supplements)	*NR3C1_a*	Cord blood	Hyper	6
van Mil, 2014	Cross-sectional	463	Netherlands	FFQ (supplements)	*5HTT_b*	Cord blood	Hypo	6
Steegers-Theunissen, 2009	Cross-sectional	120	Netherlands	Questionnaire (supplements)	*IGF2*	Whole blood	Hyper	5
Haggarty, 2013	Cross-sectional	913	UK	FFQ	*IGF2*	Cord blood	Hyper	5
Haggarty, 2013	Cross-sectional	913	UK	FFQ	*PEG3*	Cord blood	Hypo	5
Haggarty, 2013	Cross-sectional	913	UK	FFQ	*SNRPN*	Cord blood	No	5
Haggarty, 2013	Cross-sectional	913	UK	RBC	*IGF2*	Cord blood	No	6
Haggarty, 2013	Cross-sectional	913	UK	RBC	*PEG3*	Cord blood	No	6
Haggarty, 2013	Cross-sectional	913	UK	RBC	*SNRPN*	Cord blood	No	6
Hoyo, 2014	Cross-sectional	496	USA	RBC	*MEG3 (Promoter)*	Cord blood	Hypo	7
Hoyo, 2014	Cross-sectional	496	USA	RBC	*NNAT (Promoter)*	Cord blood	Hypo	7
Hoyo, 2014	Cross-sectional	496	USA	RBC	*PEG10/SGCE (Promoter)*	Cord blood	Hypo	7

First author, year	Study design	N	Country	Exposure assessment	Genes	Methylation Tissue	Direction results	QS
Hoyo, 2014	Cross-sectional	496	USA	RBC	*MEG3-IG (Promoter)*	Cord blood	Hypo	7
Hoyo, 2014	Cross-sectional	496	USA	RBC	*PLAGL1 (Promoter)*	Cord blood	Hypo	7
Hoyo, 2014	Cross-sectional	496	USA	RBC	*PEG3 (Promoter)*	Cord blood	Hypo	7
Hoyo, 2014	Cross-sectional	496	USA	RBC	*PEG1/MEST (Promoter)*	Cord blood	Hyper	7
Hoyo, 2014	Cross-sectional	496	USA	RBC	*H19*	Cord blood	Hypo	7
Hoyo, 2014	Cross-sectional	496	USA	RBC	*IGF2 (Promoter)*	Cord blood	Hyper	7
Ba, 2011	Cross-sectional	99	China	Serum	*IGF2 (Promoter)*	Cord blood	No	6
van Mil, 2014	Cross-sectional	463	Netherlands	Venous blood	*NR3C1_a*	Cord blood	Hyper	7
GENOME-WIDE								
Gonseth, 2015	Cross-sectional	347	USA	FFQ	NA	Neonatal blood	Hyper/Hypo	6
Amarasekera, 2014	Cross-sectional	23	Australia	Serum	NA	Cord blood	Hyper/Hypo	4
Joubert, 2016	Cross-sectional	1988	Netherlands, Norway	Plasma	NA	Cord blood	Hyper/Hypo	8
Mozhui, 2015	Cross-sectional	200	USA	Plasma/Serum	NA	Cord blood (from buffy coats)	No	6
CHILDHOOD								
GLOBAL								
Fryer, 2009	Cross-sectional	24	UK	Cord serum	NA	Cord blood	No	4
Haggarty, 2013	Cross-sectional	913	UK	RBC	NA	Cord blood	Hypo	5
Perng, 2012	Cross-sectional	568	Colombia	RBC	NA	Blood	No	4
GENE-SPECIFIC								
Haggarty, 2013	Cross-sectional	913	UK	RBC	*IGF2*	Cord blood	Hyper	6
Haggarty, 2013	Cross-sectional	913	UK	RBC	*PEG3*	Cord blood	Hypo	6
Haggarty, 2013	Cross-sectional	913	UK	RBC	*SNRPN*	Cord blood	No	6
ADULTHOOD								
GLOBAL								
Ono, 2012	Cross-sectional	384	Japan	FFQ &24h recall	NA	Blood	Hypo	5

First author, year	Study design	N	Country	Exposure assessment	Genes	Methylation Tissue	Direction results	QS
Protiva, 2011	Cross-sectional	20	USA	3d questionnaires & 24h recalls	NA	Blood	No	1
Gomes, 2012	Cross-sectional	126	Brazil	24h recall on 3 d	NA	Venous blood	No	3
Huang, 2012	Cross-sectional	493	USA	FFQ (supplements + diet + fortified)	NA	Blood	No	3
Huang, 2012	Cross-sectional	493	USA	FFQ (DFEs)	NA	Blood	No	3
Huang, 2012	Cross-sectional	493	USA	FFQ (Natural)	NA	Blood	Hyper	3
Boeke, 2012	Cross-sectional	534	USA	FFQ	NA	Blood	No	6
Perng, 2014	Cross-sectional	987	USA	FFQ	NA	Blood	No	3
Agodi, 2015	Cross-sectional	177	Italy	FFQ	NA	Whole blood	Hyper	5
Zhang 2012	Cross-sectional	165	US	FFQ	NA	WBC	No	4
Zhang 2012	Cross-sectional	165	US	FFQ	NA	WBC	Hyper	5
Zhang 2012	Cross-sectional	165	US	FFQ	NA	WBC	No	4
Zhang, 2011	Cross-sectional	149	US	FFQ	NA	Peripheral blood leukocytes	No	4
Zhang, 2011	Cross-sectional	149	US	FFQ	NA	Peripheral blood leukocytes	No	4
Zhang, 2011	Cross-sectional	149	US	FFQ	NA	Peripheral blood leukocytes	No	4
Ulrich, 2012	Cross-sectional	173	US	FFQ	NA	Lymphocytes	No	4
Figueiredo, 2009	Cross-sectional	388	North America	FFQ	NA	Colonic biopsie	NA (Quartiles)	3
Badiga, 2014	Cross-sectional	325	USA	Plasma	NA	Blood	Hypo	6
Pilsner, 2007	Cross-sectional	294	Bangladesh	Plasma	NA	Blood	Hyper	6
Piyathilake, 2013	Cross-sectional	470	USA	Plasma	NA	Blood	Hyper	4
Gadgil, 2014	Cross-sectional	49	India	Plasma	NA	Blood	No	3
Narayanan, 2004	Cross-sectional	408	Scotland	Plasma	NA	Venous blood	No	4
Tapp, 2013	Cross-sectional	185	UK	Plasma	NA	Rectal mucosa	No	5
Llanos, 2015	Cross-sectional	121	USA	Plasma	NA	Breast tissue	No	6
Figueiredo, 2009	Cross-sectional	388	North America	Plasma	NA	Colonic biopsie	NA (Quartiles)	4
Hanks, 2013	Cross-sectional	336	UK	Serum	NA	Colonic tissue biopsies	No	6

First author, year	Study design	N	Country	Exposure assessment	Genes	Methylation Tissue	Direction results	QS
Pufulete, 2005	Cross-sectional	68	UK	Serum	NA	Colonic mucosa	No	5
Ulrich, 2012	Cross-sectional	173	US	Serum	NA	Lymphocytes	No	5
Wang, 2012	Cross-sectional	115	China	Serum	NA	Whole blood	Hyper	6
Fenech, 1998	Cross-sectional	106	Australia	RBC	NA	Blood	No	6
Bae, 2014	Cross-sectional	408	USA	RBC (Prefortification period)	NA	Blood	Hyper	6
Bae, 2014	Cross-sectional	408	USA	RBC (Postfortification period)	NA	Blood	Hyper	6
Figueiredo, 2009	Cross-sectional	388	North America	RBC	NA	Colonic biopsie	NA (Quartiles)	4
Hanks, 2013	Cross-sectional	336	UK	RBC	NA	Colonic tissue biopsies	No	6
Pufulete, 2005	Cross-sectional	68	UK	RBC	NA	Colonic mucosa	No	5
Tapp, 2013	Cross-sectional	185	UK	RBC	NA	Rectal mucosa	No	5
Friso, 2002	Cross-sectional	292	Italy	Venous blood	NA	Whole blood	Hyper	6
Friso, 2005	Cross-sectional	198	Italy	Venous blood	NA	Whole blood	Hyper	5
Hanks, 2013	Cross-sectional	336	UK	Colonic tissue	NA	Colonic tissue biopsies	No	6
Llanos, 2015	Cross-sectional	121	USA	Breast tissue	NA	Breast tissue	No	6
Jacob, 1998	Intervention[N]	8	USA	Restricted diet	NA	Blood	Hypo	2
Rampersaud, 2000	Intervention[N]	33	USA	Diet	NA	Leukocytes	Hyper	3
Hübner, 2013	Intervention[N]	50	Germany	Supplementation	NA	Blood	No	3
Axume, 2007	Intervention[N]	43	USA	Folate depletion followed by repletion[#]	NA	Blood	Hypo	5
Aarabi, 2015	Intervention[N]	28	Canada	Folic acid (5mg/day) for 6 m	NA	Sperm	Hypo	4
Shelnutt, 2004	Intervention[N]	41	USA	Folate depletion followed by repletion[#]	NA	Leukocytes	Hyper	3
Jung, 2011	Intervention[D]	216	Netherlands	FFQ; FA (0.8 mg/d) or placebo	NA	Blood	No	8
Crider, 2011	Intervention[D]	1108	China	FA supplementation for 6m	NA	Blood[T]	Hypo	8
Basten, 2006	Intervention[D]	61	UK	Plasma	NA	Whole blood	No	5
Basten, 2006	Intervention[D]	61	UK	RBC	NA	Whole blood	No	5
Basten, 2006	Intervention[D]	61	UK	Lymphocytes	NA	Whole blood	No	5
GENE-SPECIFIC								

First author, year	Study design	N	Country	Exposure assessment	Genes	Methylation Tissue	Direction results	QS
Bollati, 2014	Cross-sectional	165	Italy	FFQ	*CD14 (Promoter)*	Blood	No	6
Bollati, 2014	Cross-sectional	165	Italy	FFQ	*Et-1 (Promoter)*	Blood	No	6
Bollati, 2014	Cross-sectional	165	Italy	FFQ	*HERV-w (Promoter)*	Blood	No	6
Bollati, 2014	Cross-sectional	165	Italy	FFQ	*iNOS (Promoter)*	Blood	No	6
Bollati, 2014	Cross-sectional	165	Italy	FFQ	*TNFα (Promoter)*	Blood	Hypo	6
Zhong, 2015	Cross-sectional	573	US	FFQ	*TLR2* (Promoter)	Whole blood	No	6
Zhang 2012	Cross-sectional	165	US	FFQ	*IL-6* (Promoter)	WBC	No	5
Zhang 2012	Cross-sectional	165	US	FFQ	*IL-6* (Promoter)	WBC	No	5
Zhang 2012	Cross-sectional	165	US	FFQ	*IL-6* (Promoter)	WBC	No	5
Stidley, 2010	Cross-sectional	1101	USA	FFQ	Index*	Sputum	Hypo	6
Dhillon, 2007	Cross-sectional	379	India	Questionnaire	*GSTM1* (Promoter)	Semen	Hypo	5
Ottini, 2015	Cross-sectional	21	Italy	Plasma	Index**	Blood	Hypo	5
Piyathilake, 2014	Cross-sectional	315	USA	Plasma	*HPV 16* (Promoter)	Blood	Hyper	5
Llanos, 2015	Cross-sectional	138	USA	Plasma	$p16^{INK4a}$ (Promoter)	Breast tissue	Hypo	7
Tapp, 2013	Cross-sectional	185	UK	Plasma	Index***	Rectal mucosa	Hyper	6
Tapp, 2013	Cross-sectional	185	UK	Plasma	*HPP1*	Rectal mucosa	Hyper	6
Tapp, 2013	Cross-sectional	185	UK	Plasma	*APC*	Rectal mucosa	No	6
Tapp, 2013	Cross-sectional	185	UK	Plasma	*SFRP1*	Rectal mucosa	Hyper	6
Tapp, 2013	Cross-sectional	185	UK	Plasma	*SFRP2*	Rectal mucosa	Hyper	6
Tapp, 2013	Cross-sectional	185	UK	Plasma	*SOX17*	Rectal mucosa	No	6
Tapp, 2013	Cross-sectional	185	UK	Plasma	*WIF1*	Rectal mucosa	Hyper	6
Tapp, 2013	Cross-sectional	185	UK	Plasma	*ESR1*	Rectal mucosa	No	6
Tapp, 2013	Cross-sectional	185	UK	Plasma	*MYOD*	Rectal mucosa	Hyper	6
Tapp, 2013	Cross-sectional	185	UK	Plasma	*N33*	Rectal mucosa	No	6
Hanks, 2013	Cross-sectional	336	UK	Serum	*MLH1* (Promoter)	Colonic tissue biopsies	No	7
Hanks, 2013	Cross-sectional	336	UK	Serum	*MGMT* (Promoter)	Colonic tissue biopsies	Hypo	7
Hanks, 2013	Cross-sectional	336	UK	Serum	*APC* (Promoter)	Colonic tissue biopsies	No	7

First author, year	Study design	N	Country	Exposure assessment	Genes	Methylation Tissue	Direction results	QS
Hanks, 2013	Cross-sectional	336	UK	Serum	*IGF2* (Promoter)	Colonic tissue biopsies	No	7
Hanks, 2013	Cross-sectional	336	UK	Serum	*MYOD1* (Promoter)	Colonic tissue biopsies	No	7
Hanks, 2013	Cross-sectional	336	UK	Serum	*N33* (Promoter)	Colonic tissue biopsies	No	7
Hanks, 2013	Cross-sectional	336	UK	Serum	*ESR1* (Promoter)	Colonic tissue biopsies	No	7
Hanks, 2013	Cross-sectional	336	UK	RBC	*MLH1* (Promoter)	Colonic tissue biopsies	No	7
Hanks, 2013	Cross-sectional	336	UK	RBC	*MGMT* (Promoter)	Colonic tissue biopsies	No	7
Hanks, 2013	Cross-sectional	336	UK	RBC	*APC* (Promoter)	Colonic tissue biopsies	No	7
Hanks, 2013	Cross-sectional	336	UK	RBC	*IGF2* (Promoter)	Colonic tissue biopsies	No	7
Hanks, 2013	Cross-sectional	336	UK	RBC	*MYOD1* (Promoter)	Colonic tissue biopsies	No	7
Hanks, 2013	Cross-sectional	336	UK	RBC	*N33* (Promoter)	Colonic tissue biopsies	No	7
Hanks, 2013	Cross-sectional	336	UK	RBC	*ESR1* (Promoter)	Colonic tissue biopsies	No	7
Tapp, 2013	Cross-sectional	185	UK	RBC	Index***	Rectal mucosa	Hyper	6
Tapp, 2013	Cross-sectional	185	UK	RBC	*HPP1*	Rectal mucosa	No	6
Tapp, 2013	Cross-sectional	185	UK	RBC	*APC*	Rectal mucosa	No	6
Tapp, 2013	Cross-sectional	185	UK	RBC	*SFRP1*	Rectal mucosa	Hyper	6
Tapp, 2013	Cross-sectional	185	UK	RBC	*SFRP2*	Rectal mucosa	Hyper	6
Tapp, 2013	Cross-sectional	185	UK	RBC	*SOX17*	Rectal mucosa	No	6
Tapp, 2013	Cross-sectional	185	UK	RBC	*WIF1*	Rectal mucosa	No	6
Tapp, 2013	Cross-sectional	185	UK	RBC	*ESR1*	Rectal mucosa	No	6
Tapp, 2013	Cross-sectional	185	UK	RBC	*MYOD*	Rectal mucosa	Hyper	6
Tapp, 2013	Cross-sectional	185	UK	RBC	*N33*	Rectal mucosa	No	6
Hanks, 2013	Cross-sectional	336	UK	Colonic tissue	*MLH1* (Promoter)	Colonic tissue biopsies	No	7
Hanks, 2013	Cross-sectional	336	UK	Colonic tissue	*MGMT* (Promoter)	Colonic tissue biopsies	No	7
Hanks, 2013	Cross-sectional	336	UK	Colonic tissue	*APC* (Promoter)	Colonic tissue biopsies	No	7
Hanks, 2013	Cross-sectional	336	UK	Colonic tissue	*IGF2* (Promoter)	Colonic tissue biopsies	No	7
Hanks, 2013	Cross-sectional	336	UK	Colonic tissue	*MYOD1* (Promoter)	Colonic tissue biopsies	No	7
Hanks, 2013	Cross-sectional	336	UK	Colonic tissue	*N33* (Promoter)	Colonic tissue biopsies	No	7

First author, year	Study design	N	Country	Exposure assessment	Genes	Methylation Tissue	Direction results	QS
Hanks, 2013	Cross-sectional	336	UK	Colonic tissue	*ESR1* (Promoter)	Colonic tissue biopsies	No	7
Llanos, 2015	Cross-sectional	138	USA	Breast tissue	$p16^{INK4a}$ (Promoter)	Breast tissue	Hypo	7
Aarabi, 2015	Intervention[N]	30	Canada	FA (5mg/d) for 6m	*H19*	Sperm	No	5
Aarabi, 2015	Intervention[N]	30	Canada	FA (5mg/d) for 6m	*DLK1/GTL2*	Sperm	No	5
Aarabi, 2015	Intervention[N]	30	Canada	FA (5mg/d) for 6m	*MEST*	Sperm	No	5
Aarabi, 2015	Intervention[N]	30	Canada	FA (5mg/d) for 6m	*SNRPN*	Sperm	No	5
Aarabi, 2015	Intervention[N]	30	Canada	FA (5mg/d) for 6m	*PLAGL1*	Sperm	No	5
Aarabi, 2015	Intervention[N]	30	Canada	FA (5mg/d) for 6m	*KCNQ1OT1*	Sperm	No	5
Al-Ghnaniem Abbadi, 2013	Intervention[D]	29	UK	FA (400 µg/d) or placebo for 10w	*ESR1* (Promoter)	Colonic mucosa[T]	No	7
Al-Ghnaniem Abbadi, 2013	Intervention[D]	29	UK	FA (400 µg/d) or placebo for 10w	*MLH1* (Promoter)	Colonic mucosa[T]	No	7
GENOME-WIDE								
Song, 2016	Cross-sectional	81	USA	Breast tissue	NA	Breast tissue	Hyper/Hypo	4

[N] Non-blinded, [D] Double-blinded, FA = Folic Acid

[T] Tissues: Brain, skin, heart, kidney, lung & liver; Colonic mucosa (Normal-appearing); Blood (Coagulated & uncoagulated)

[#]Folate depletion (7w, 115 or 135 µg DFE/d) followed by repletion (7w, 400 or 800 µg DFE/d)

*Index *(p16, MGMT, DAPK, RASSF1A, PAX5α,PAX5β, GATA4 & GATA5)* (Promoter)

**Index *(p16, FHIT, RAR, CDH1, DAPK1, hTERT, RASSF1A, MGMT, BRCA1 & PALB2)* (Promoter)

***Index *(HPP1, APC, SFRP1, SFRP2, SOX17, WIF1, ESR1, MYOD & N33)*

Table 4. Study characteristics and results: Vitamin B1

First author, year	Study design	N	Country	Exposure assessment	Genes	Tissue	Direction results	QS
ADULTHOOD								
GLOBAL								
Marques-Rocha, 2016	Cross-sectional	156	Brazil	72h food record	NA	WBCs	Hypo	5

Table 5. Study characteristics and results: Vitamin B2

First author, year	Study design	N	Country	Exposure assessment	Genes	Tissue	Direction results	QS
MATERNAL OFFSPRING RELATIONSHIP								
GENE-SPECIFIC								
Azzi, 2014	Cross-sectional	2002	France	FFQ & 24h recall	*ZAC1*	Cord blood	Hyper	6
MATERNAL OFFSPRING RELATIONSHIP								
GENE-SPECIFIC								
Marques-Rocha, 2016	Cross-sectional	156	Brazil	72h food record	NA	WBCs	Hyper	5
Ono, 2012	Cross-sectional	384	Japan	FFQ & 24h recall	NA	Blood	No	5
Zhang 2012	Cross-sectional	165	USA	FFQ	NA	WBCs	No	4
Zhang, 2011	Cross-sectional	149	USA	FFQ	NA	Peripheral blood leukocytes	No	4
Figueiredo, 2009	Cross-sectional	388	North America	FFQ	NA	Biopsies from right and left colon	NA (Quartiles)	3
Figueiredo, 2009	Cross-sectional	388	North America	Plasma	NA	Biopsies from right and left colon	NA (Quartiles)	4

Table 6. Study characteristics and results: Vitamin B3

First author, year	Study design	N	Country	Exposure assessment	Genes	Tissue	Direction results	QS
MATERNAL OFFSPRING RELATIONSHIP								
GENE-SPECIFIC								
Azzi, 2014	Cross-sectional	2002	France	FFQ & 24h recall	*ZAC1*	Cord blood	No	6
ADULTHOOD								
GLOBAL								
Marques-Rocha, 2016	Cross-sectional	156	Brazil	72h food record	NA	WBCs	Hypo	5

Table 7. Study characteristics and results: Vitamin B6

First author, year	Study design	N	Country	Exposure assessment	Genes	Tissue	Direction results	QS
MATERNAL OFFSPRING RELATIONSHIP								
GENE-SPECIFIC								
Azzi, 2014	Cross-sectional	2002	France	FFQ & 24h recall	*ZAC1*	Cord blood	No	6
ADULTHOOD								
GLOBAL								
Gomes, 2012	Cross-sectional	126	Brazil	24h recall on 3 different days	NA	Venous blood	Hypo	3
Ono, 2012	Cross-sectional	384	Japan	FFQ & 24h recall	NA	Blood	No	5
Huang, 2012	Cross-sectional	493	USA	FFQ	NA	Blood	No	3
Perng, 2014	Cross-sectional	987	USA	FFQ	NA	Blood	No	3
Zhang, 2011	Cross-sectional	149	USA	FFQ	NA	Peripheral blood leukocytes	No	4
Ulrich, 2012	Cross-sectional	173	USA	FFQ	NA	Lymphocytes	No	4
Zhang 2012	Cross-sectional	165	USA	FFQ	NA	WBCs	No	4
Figueiredo, 2009	Cross-sectional	388	North America	FFQ	NA	Biopsies from right and left colon	NA (Quartiles)	3
Figueiredo, 2009	Cross-sectional	388	North America	Plasma	NA	Biopsies from right and left colon	NA (Quartiles)	4
Friso, 2002	Cross-sectional	292	Italy	Venous blood	NA	Whole blood	No	6
Hübner, 2013	Intervention (Non-blinded)	50	Germany	Supplements	NA	Blood	No	3
GENE-SPECIFIC								
Zhang 2012	Cross-sectional	165	USA	FFQ	*IL-6*	WBCs	No	5

Table 8. Study characteristics and results: Vitamin B12

First author, year	Study design	N	Country	Exposure assessment	Genes	Tissue	Direction results	QS
MATERNAL OFFSPRING RELATIONSHIP								
GLOBAL								
Boeke, 2012	Cross-sectional	534	USA	FFQ	NA	Cord blood	No	6
McKay, 2012	Cross-sectional	158	UK	Serum	NA	Cord blood	Hyper	5
GENE-SPECIFIC								
Azzi, 2014	Cross-sectional	2002	France	FFQ & 24h recall	*ZAC1*	Cord blood	Hyper	6
Ba, 2011	Cross-sectional	99	China	Serum	*IGF2*	Cord blood	Hypo	6
CHILDHOOD								
GLOBAL								
Perng, 2012	Cross-sectional	568	Colombia	Plasma	*NA*	Blood	No	4
GENE-SPECIFIC								
McKay, 2012	Cross-sectional	158	UK	Serum cord blood	*IGFBP3*	Cord blood	Hypo	7
ADULTHOOD								
GLOBAL								
Gomes, 2012	Cross-sectional	126	Brazil	24h recall (3 d)	NA	Venous blood	No	3
Ono, 2012	Cross-sectional	384	Japan	FFQ &24 h recall	NA	Blood	No	5
Huang, 2012	Cross-sectional	493	USA	FFQ	NA	Blood	No	3
Perng, 2014	Cross-sectional	987	USA	FFQ	NA	Blood	No	5
Boeke, 2012	Cross-sectional	534	USA	FFQ	NA	Blood	No	6
Ulrich, 2012	Cross-sectional	173	USA	FFQ	NA	Lymphocytes	No	4
Zhang 2012	Cross-sectional	165	USA	FFQ	NA	WBCs	No	4
Zhang, 2011	Cross-sectional	149	USA	FFQ	NA	Peripheral blood leukocytes	No	4
Figueiredo, 2009	Cross-sectional	388	North America	Plasma	NA	Biopsie from right and left colon	NA (Quartiles)	4
Narayanan, 2004	Cross-sectional	408	Scotland	Plasma	NA	Venous blood	No	4
Quinlivan, 2013	Cross-sectional	376	China	Plasma	NA	Coagulated Blood clots vs. uncoagulated EDTA-Blood cells	Hyper	4
Quinlivan, 2013	Cross-sectional	376	China	Plasma	NA	Blood clots	Hyper	6
Gadgil, 2014	Cross-sectional	49	India	Plasma	NA	Blood	No	3

First author, year	Study design	N	Country	Exposure assessment	Genes	Tissue	Direction results	QS
Tapp, 2013	Cross-sectional	185	UK	Plasma	NA	Rectal mucosa	No	5
Fenech, 1998	Cross-sectional	106	Australia	Serum	NA	Blood	No	6
Friso, 2002	Cross-sectional	292	Italy	venous blood	NA	Whole blood	No	6
Ulrich, 2012	Cross-sectional	173	USA	Serum	NA	Lymphocytes	No	5
Pufulete, 2005	Cross-sectional	68	UK	Serum	NA	Colonic mucosa	No	3
Hübner, 2013	Intervention (Non-blinded)	50	Germany	Supplements	NA	Blood	No	3
GENE-SPECIFIC								
Bollati, 2014	Cross-sectional	165	Italy	FFQ	*CD14*	Blood	No	6
Bollati, 2014	Cross-sectional	165	Italy	FFQ	*Et-1*	Blood	No	6
Bollati, 2014	Cross-sectional	165	Italy	FFQ	*HERV-w*	Blood	No	6
Bollati, 2014	Cross-sectional	165	Italy	FFQ	*iNOS*	Blood	No	6
Bollati, 2014	Cross-sectional	165	Italy	FFQ	*TNFα*	Blood	No	6
Zhang 2012	Cross-sectional	165	USA	FFQ	*IL-6*	WBCs	No	5
Zhong, 2015	Cross-sectional	573	USA	FFQ	*TLR2*	Whole blood	No	6
Piyathilake, 2014	Cross-sectional	315	USA	Plasma	*HPV 16*	Blood	Hyper	5
Ottini, 2015	Cross-sectional	22	Italy	Plasma	Index*	Blood	No	3
Tapp, 2013	Cross-sectional	185	UK	Plasma	Index**	Rectal mucosa	No	6
Tapp, 2013	Cross-sectional	185	UK	Plasma	*HPP1*	Rectal mucosa	No	6
Tapp, 2013	Cross-sectional	185	UK	Plasma	*APC*	Rectal mucosa	No	6
Tapp, 2013	Cross-sectional	185	UK	Plasma	*SFRP1*	Rectal mucosa	No	6
Tapp, 2013	Cross-sectional	185	UK	Plasma	*SFRP2*	Rectal mucosa	No	6
Tapp, 2013	Cross-sectional	185	UK	Plasma	*SOX17*	Rectal mucosa	No	6
Tapp, 2013	Cross-sectional	185	UK	Plasma	*WIF1*	Rectal mucosa	No	6
Tapp, 2013	Cross-sectional	185	UK	Plasma	*ESR1*	Rectal mucosa	No	6
Tapp, 2013	Cross-sectional	185	UK	Plasma	*MYOD*	Rectal mucosa	No	6
Tapp, 2013	Cross-sectional	185	UK	Plasma	*N33*	Rectal mucosa	No	6
Al-Ghnaniem,2007	Cross-sectional	73	UK	Venous blood	*ERα*	Normal-appearing colonic mucosa	Hypo	6

*Index (*p16, FHIT, RAR, CDH1, DAPK1, hTERT, RASSF1A, MGMT, BRCA1 & PALB2), ***Index (*HPP1, APC, SFRP1, SFRP2, SOX17, WIF1, ESR1, MYOD & N33*)

Table 9. Study characteristics and results: Vitamin C

First author, year	Study design	N	Country	Exposure assessment	Genes	Tissue	Direction results	QS
ADULTHOOD								
GLOBAL								
Piyathilake, 2011	Cross-sectional	376	USA	Plasma	NA	PBMC	No	6
Piyathilake, 2011	Cross-sectional	376	USA	Plasma	NA	Cervical cells	No	6
GENE-SPECIFIC								
Bollati, 2014	Cross-sectional	165	Italy	FFQ	CD14	Blood	No	6
Bollati, 2014	Cross-sectional	165	Italy	FFQ	Et-1	Blood	No	6
Bollati, 2014	Cross-sectional	165	Italy	FFQ	HERV-w	Blood	No	6
Bollati, 2014	Cross-sectional	165	Italy	FFQ	iNOS	Blood	No	6
Bollati, 2014	Cross-sectional	165	Italy	FFQ	TNFα	Blood	No	6
Stidley, 2010	Cross-sectional	1101	USA	FFQ	Index (p16, MGMT, DAPK, RASSF1A, PAX5α, PAX5β, GATA4, and GATA5)	Sputum	No	6
Piyathilake, 2014	Cross-sectional	315	USA	Plasma	HPV 16	Blood	No	5
de la Iglesia, 2014	Intervention[N]	47	Spain	48 h weighed food record	PON1	Venous blood	Hypo	6

[N] Non-blinded

Table 10. Study characteristics and results: Vitamin D

First author, year	Study design	N	Country	Exposure assessment	Genes	Tissue	Direction results	QS
MATERNAL OFFSPRING RELATIONSHIP								
GENE-SPECIFIC								
Novakovic, 2016	Cross-sectional	68	Canada	Serum	*CYP24A1*	Placenta	No	4
Harvey, 2014	Cross-sectional	230	UK	Plasma	*RXRA*	umbilical cord	Hypo	5
GENOME-WIDE								
Mozhui, 2015	Cross-sectional	147	USA	Plasma/Serum	NA	Cord blood from buffy coats	No	6
CHILDHOOD								
GLOBAL								
Zhu, 2016	Cross-sectional	454	USA	Plasma	NA	Peripheral Blood	Hyper	6
Zhu, 2016	Intervention[D]	58	USA	Supplements	NA	Peripheral Blood	Hyper	7
GENE-SPECIFIC								
Zhu, 2013	Cross-sectional	22	USA	Plasma	*DHCR7*	Whole blood leukocytes	Opposite	4
Zhu, 2013	Cross-sectional	22	USA	Plasma	*CYP2R1*	Whole blood leukocytes	Hypo	4
Zhu, 2013	Cross-sectional	22	USA	Plasma	*CYP24A1*	Whole blood leukocytes	Hyper	4
Novakovic, 2016	Cross-sectional	68	Canada	Serum	*CYP24A1*	Placenta	No	4
GENOME-WIDE								
Zhu, 2013	Cross-sectional	22	USA	Plasma	NA	Whole blood leukocytes	Hyper/Hypo	5
ADULTHOOD								
GLOBAL								
Tapp, 2013	Cross-sectional	185	UK	Plasma	NA	Rectal mucosa	No	5
Nair-Shalliker, 2014	Cross-sectional	208	Australia	Serum	NA	Lymphocytes	No	5
Hübner, 2013	Intervention[N]	50	Germany	Supplements	NA	Blood	No	3
GENE-SPECIFIC								
Bollati, 2014	Cross-sectional	165	Italy	FFQ	*CD14*	Blood	No	6
Bollati, 2014	Cross-sectional	165	Italy	FFQ	*Et-1*	Blood	No	6

First author, year	Study design	N	Country	Exposure assessment	Genes	Tissue	Direction results	QS
Bollati, 2014	Cross-sectional	165	Italy	FFQ	*HERV-w*	Blood	No	6
Bollati, 2014	Cross-sectional	165	Italy	FFQ	*iNOS*	Blood	No	6
Bollati, 2014	Cross-sectional	165	Italy	FFQ	*TNFα*	Blood	No	6
Ashktorab, 2011	Cross-sectional	187	USA	Serum	*DKK1*	Whole blood	No	7
Tapp, 2013	Cross-sectional	185	UK	Plasma	Index*	Rectal mucosa	Hypo	6
Tapp, 2013	Cross-sectional	185	UK	Plasma	*HPP1*	Rectal mucosa	No	6
Tapp, 2013	Cross-sectional	185	UK	Plasma	*APC*	Rectal mucosa	Hypo	6
Tapp, 2013	Cross-sectional	185	UK	Plasma	*SFRP1*	Rectal mucosa	No	6
Tapp, 2013	Cross-sectional	185	UK	Plasma	*SFRP2*	Rectal mucosa	No	6
Tapp, 2013	Cross-sectional	185	UK	Plasma	*SOX17*	Rectal mucosa	No	6
Tapp, 2013	Cross-sectional	185	UK	Plasma	*WIF1*	Rectal mucosa	Hypo	6
Tapp, 2013	Cross-sectional	185	UK	Plasma	*ESR1*	Rectal mucosa	No	6
Tapp, 2013	Cross-sectional	185	UK	Plasma	*MYOD*	Rectal mucosa	Hypo	6
Tapp, 2013	Cross-sectional	185	UK	Plasma	*N33*	Rectal mucosa	No	6

[N] Non-blinded

[D] Double-blinded

*Index (*HPP1, APC, SFRP1, SFRP2, SOX17, WIF1, ESR1, MYOD, N33*)

Table 11. Study characteristics and results: Vitamin E

First author, year	Study design	N	Country	Exposure assessment	Genes	Tissue	Direction results	QS
ADULTHOOD								
GLOBAL								
Piyathilake, 2011	Cross-sectional	376	USA	Plasma	NA	PBMC	Hypo	6
Piyathilake, 2011	Cross-sectional	376	USA	Plasma	NA	Cervical cells	No	6
GENE-SPECIFIC								
Bollati, 2014	Cross-sectional	165	Italy	FFQ	*CD14*	Blood	No	6
Bollati, 2014	Cross-sectional	165	Italy	FFQ	*Et-1*	Blood	No	6
Bollati, 2014	Cross-sectional	165	Italy	FFQ	*HERV-w*	Blood	No	6
Bollati, 2014	Cross-sectional	165	Italy	FFQ	*iNOS*	Blood	No	6
Bollati, 2014	Cross-sectional	165	Italy	FFQ	*TNFα*	Blood	No	6
de la Iglesia, 2014	Intervention[N]	47	Spain	48 h weighed food record	*PON1*	Venous blood	Hypo	6
Stidley, 2010	Cross-sectional	1101	USA	FFQ	Index*	Sputum	No	6

[N] Non-blinded

*Index (p16, MGMT, DAPK, RASSF1A, PAX5α, PAX5β, GATA4 & GATA5)

Table 12. Study characteristics and results: Choline and betaine

Nutritional exposure	First author, year	Study design	N	Country	Exposure assessment	Genes	Tissue	Direction results	QS
MATERNAL OFFSPRING RELATIONSHIP									
GLOBAL									
Choline	Boeke, 2012	Cross-sectional	534	USA	FFQ	NA	Cord blood	Hypo	6
Choline	Jiang, 2012	Intervention[N]	24	USA	12-wk controlled feeding study	NA	Placenta	Hyper	4
Choline	Jiang, 2012	Intervention[N]	24	USA	12-wk controlled feeding study	NA	Cord blood	No	4
Betaine	Boeke, 2012	Cross-sectional	534	USA	FFQ	NA	Cord blood	Hypo	6
GENE-SPECIFIC									

Nutritional exposure	First author, year	Study design	N	Country	Exposure assessment	Genes	Tissue	Direction results	QS
Choline	Jiang, 2012	Intervention[N]	24	USA	12-wk controlled feeding study	*CRH*	Cord blood	Hypo	5
Choline	Jiang, 2012	Intervention[N]	24	USA	12-wk controlled feeding study	*NR3C1*	Cord blood	Hypo	5
Choline	Jiang, 2012	Intervention[N]	24	USA	12-wk controlled feeding study	*GNAS-AS1*	Cord blood	No	5
Choline	Jiang, 2012	Intervention[N]	24	USA	12-wk controlled feeding study	*IGF2*	Cord blood	No	5
Choline	Jiang, 2012	Intervention[N]	24	USA	12-wk controlled feeding study	*IL10*	Cord blood	No	5
Choline	Jiang, 2012	Intervention[N]	24	USA	12-wk controlled feeding study	*LEP*	Cord blood	No	5
Choline	Jiang, 2012	Intervention[N]	24	USA	12-wk controlled feeding study	*CRH*	Placenta	Hyper	5
Choline	Jiang, 2012	Intervention[N]	24	USA	12-wk controlled feeding study	*NR3C1*	Placenta	Hyper	5
Choline	Jiang, 2012	Intervention[N]	24	USA	12-wk controlled feeding study	*GNAS-AS1*	Placenta	No	5
Choline	Jiang, 2012	Intervention[N]	24	USA	12-wk controlled feeding study	*IGF2*	Placenta	No	5
Choline	Jiang, 2012	Intervention[N]	24	USA	12-wk controlled feeding study	*IL10*	Placenta	No	5
Choline	Jiang, 2012	Intervention[N]	24	USA	12-wk controlled feeding study	*LEP*	Placenta	No	5
ADULTHOOD									
GLOBAL									
Choline	Boeke, 2012	Cross-sectional	534	USA	FFQ	NA	Blood	No	6
Choline	Jiang, 2012	Intervention[N]	24	USA	12-wk controlled feeding study	NA	Blood	No	4
Choline	Shin, 2010	Intervention[N]	60	USA	Intervention*	NA	Leukocytes	Hyper	3
Betaine	Boeke, 2012	Cross-sectional	534	USA	FFQ	NA	Blood	No	6
GENE-SPECIFIC									
Choline	Jiang, 2012	Intervention[N]	24	USA	12-wk controlled feeding study	*CRH*	Blood	No	5
Choline	Jiang, 2012	Intervention[N]	24	USA	12-wk controlled feeding study	*NR3C1*	Blood	No	5

[N] Non-blinded

*Intervention (12-wk of 300, 550, 1100, or 2200 mg/d)

Table 13. Study characteristics and results: Minerals and trace elements

Nutritional exposure	First author, year	Study design	N	Country	Exposure assessment	Genes	Tissue	Direction results	QS
CHILDHOOD									
GLOBAL									
Iron (Ferritin)	Perng, 2012	Cross-sectional	568	Colombia	Plasma	NA	Blood	No	4
Zinc	Perng, 2012	Cross-sectional	568	Colombia	Serum	NA	Blood	No	4
ADULTHOOD									
GLOBAL									
Calcium	Hübner, 2013	Intervention[N]	50	Germany	Supplements	NA	Blood	No	3
Copper	Marques-Rocha, 2016	Cross-sectional	156	Brazil	72h food record	NA	WBCs	Hyper	5
Iron	Marques-Rocha, 2016	Cross-sectional	156	Brazil	72h food record	NA	WBCs	Hyper	5
Magnesium	Gomes, 2012	Cross-sectional	126	Brazil	24h recall o(3 d)	NA	Venous blood	Hypo	3
Phosphate	McChelland, 2016	Cross-sectional	666	UK	Serum	NA	serum Blood	Hypo	6
Selenium	Tapp, 2013	Cross-sectional	185	UK	Plasma	NA	Rectal mucosa	Hyper	5
Zinc	Perng, 2014	Cross-sectional	987	USA	FFQ	NA	Blood	No	5
Zinc	Gomes, 2012	Cross-sectional	126	Brazil	24h recall (3 d)	NA	Venous blood	No	3
GENE-SPECIFIC									
Sodium (Salt)	Shimazu, T2015	Cross-sectional	281	Japan	FFQ	*miR-124a-3*	Gastric mucosa	No	6
Sodium (Salt)	Shimazu, T2015	Cross-sectional	281	Japan	FFQ	*EMX1*	Gastric mucosa	No	6
Sodium (Salt)	Shimazu, T2015	Cross-sectional	281	Japan	FFQ	*NKX6-1*	Gastric mucosa	No	6
Selenium	Tapp, 2013	Cross-sectional	185	UK	Plasma	*HPP1*	Rectal mucosa	No	6
Selenium	Tapp, 2013	Cross-sectional	185	UK	Plasma	*APC*	Rectal mucosa	No	6
Selenium	Tapp, 2013	Cross-sectional	185	UK	Plasma	*SFRP1*	Rectal mucosa	No	6
Selenium	Tapp, 2013	Cross-sectional	185	UK	Plasma	*SFRP2*	Rectal mucosa	No	6
Selenium	Tapp, 2013	Cross-sectional	185	UK	Plasma	*SOX17*	Rectal mucosa	No	6
Selenium	Tapp, 2013	Cross-sectional	185	UK	Plasma	*WIF1*	Rectal mucosa	No	6
Selenium	Tapp, 2013	Cross-sectional	185	UK	Plasma	*ESR1*	Rectal mucosa	No	6
Selenium	Tapp, 2013	Cross-sectional	185	UK	Plasma	*MYOD*	Rectal mucosa	No	6
Selenium	Tapp, 2013	Cross-sectional	185	UK	Plasma	*N33*	Rectal mucosa	No	6
Selenium	Tapp, 2013	Cross-sectional	185	UK	Plasma	Index*	Rectal mucosa	No	6

[N] Non-blinded, *Index *(HPP1, APC, SFRP1, SFRP2, SOX17, WIF1, ESR1, MYOD, N33)*

Table 14. Study characteristics and results: Combined nutrients

Nutritional exposure	First author, year	Study design	N	Country	Exposure assessment	Genes	Tissue	Direction results	QS
MATERNAL OFFSPRING RELATIONSHIP									
GENOME-WIDE									
Multivitamin	Khulan, 2012	Intervention[D]	59	Gambia	Supplements	NA	Cord blood	Hyper/Hypo	6
Multivitamin	Khulan, 2012	Intervention[D]	25	Gambia	Supplements	NA	Peripheral Blood from infants (~9m old)	Hyper/Hypo	6
ADULTHOOD									
GLOBAL									
Folate & Vit B12	Piyathilake, 2011	Cross-sectional	376	USA	Plasma	NA	PBMC	Hyper	6
Folate & Vit B12	Piyathilake, 2011	Cross-sectional	376	USA	Plasma	NA	Cervical cells	No	6
Vit B6, B12, D, Folate & Calcium (Vs. Control: Vit D & Calcium)	Pusceddu, 2016	Intervention[D]	65	Germany	Supplements; Serum after 1y	NA	Blood	Hyper/Hypo	5
Vit D & Calcium	Pusceddu, 2016	Intervention[D]	65	Germany	Serum	NA	Blood	Hypo	5
Folate & Vit B12	Fenech, 1998	Intervention[D]	64	Australia	Supplements	NA	Blood	No	6
GENE-SPECIFIC									
Multivitamin	Stidley, 2010	Cross-sectional	1101	USA	FFQ	Index**	Sputum	Hypo	6
Calcium & Vit D	Zhou, 2014	Intervention[N]	446	USA	Supplements*	*CYP24A1*	Serum	Hypo	5
Calcium & Vit D	Zhou, 2014	Intervention[N]	446	USA	Supplements*	*CYP2R1*	Serum	No	5
Calcium & Vit D	Zhou, 2014	Intervention[N]	446	USA	Supplements*	*CYP27A1*	Serum	No	5
Calcium & Vit D	Zhou, 2014	Intervention[N]	446	USA	Supplements*	*CYP27B1*	Serum	No	5
GENOME-WIDE									
Folate & Vit B12	Kok, 2015	Intervention[D]	92	Netherlands	Supplements; Serum after 2y	NA	Buffy coats	Hyper/Hypo	8
Folate & Vit B12	Kok, 2015	Intervention[D]	92	Netherlands	Supplements; Serum after 2y	NA	Buffy coats	Hyper/Hypo	8

[N] Non-blinded; [D] Double-blinded; Vit: Vitamin; *Supplements (Vitamin D (1100 IU/d) & calcium (1400-1500 mg/d))
**Index *(p16, MGMT, DAPK, RASSF1A, PAX5α, PAX5β, GATA4, and GATA5)*

Table 15. Study characteristics and results: Bioactive compounds

Nutritional exposure	First author, year	Study design	N	Country	Exposure assessment	Genes	Tissue	Direction results	QS
ADULTHOOD									
GLOBAL									
Carotene	Piyathilake, 2011	Cross-sectional	376	USA	Plasma	NA	PBMC	No	6
Carotene	Piyathilake, 2011	Cross-sectional	376	USA	Plasma	NA	Cervical cells	No	6
GENE-SPECIFIC									
Carotenoids	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*CD14*	Blood	No	6
Carotenoids	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*Et-1*	Blood	No	6
Carotenoids	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*HERV-w*	Blood	Hyper	6
Carotenoids	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*iNOS*	Blood	No	6
Carotenoids	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*TNFα*	Blood	Hypo	6
Carotene	Piyathilake, 2014	Cross-sectional	315	USA	Plasma	*HPV 16*	Blood	No	5
Carotene	Stidley, 2010	Cross-sectional	1101	USA	FFQ	Index*	Sputum	No	6
Lycopene	Stidley, 2010	Cross-sectional	1101	USA	FFQ	Index*	Sputum	No	6
Lycopene	de la Iglesia, 2014	Intervention[N]	47	Spain	48h weighed food record	*PON1*	Venous blood	Hypo	6
Flavonoids	Zhong, 2015	Cross-sectional	573	USA	FFQ	*TLR2*	Whole blood	Hypo	6
Polyphenols & Flavonoids	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*CD14*	Blood	No	6
Polyphenols & Flavonoids	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*Et-1*	Blood	No	6
Polyphenols & Flavonoids	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*HERV-w*	Blood	No	6
Polyphenols & Flavonoids	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*iNOS*	Blood	No	6
Polyphenols & Flavonoids	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*TNFα*	Blood	No	6
Lutein & zeaxanthin	Stidley, 2010	Cross-sectional	1101	USA	FFQ	Index*	Sputum	No	6

[N] Non-blinded

*Index *(p16, MGMT, DAPK, RASSF1A, PAX5α, PAX5β, GATA4, and GATA5)*

Table 16. Study characteristics and results: Fat and fatty acids

Nutritional exposure	First author, year	Study design	N	Country	Exposure assessment	Genes	Tissue	Direction results	QS
MATERNAL OFFSPRING RELATIONSHIP									
GLOBAL									
Fat	Rerkasem, 2015	Cross-sectional	249	Thailand	FFQ & 24h recall	NA	Venous blood	No	3
PUFA n-3*	Lee, 2013	Intervention[D]	261	Mexico	Supplements	NA	CBMCs	Hyper	8
GENE-SPECIFIC									
Fat	Godfrey, 2011	Cross-sectional	78	UK	FFQ	*RXRA*	Cord tissue	No	3
Fat	Godfrey, 2011	Cross-sectional	78	UK	FFQ	*eNOS*	Cord tissue	No	3
PUFA n-3*	Lee, 2013	Intervention[D]	261	Mexico	Supplements	*IFNγ*	CBMCs	No	9
PUFA n-3*	Lee, 2013	Intervention[D]	261	Mexico	Supplements	*TNF-α*	CBMCs	No	9
PUFA n-3*	Lee, 2013	Intervention[D]	261	Mexico	Supplements	*IL13*	CBMCs	No	9
PUFA n-3*	Lee, 2013	Intervention[D]	261	Mexico	Supplements	*GATA3*	CBMCs	No	9
PUFA n-3*	Lee, 2013	Intervention[D]	261	Mexico	Supplements	*STAT3*	CBMCs	No	9
PUFA n-3*	Lee, 2013	Intervention[D]	261	Mexico	Supplements	*IL10*	CBMCs	No	9
PUFA n-3*	Lee, 2013	Intervention[D]	261	Mexico	Supplements	*FOXP3*	CBMCs	No	9
PUFA n-3*	Lee, 2014	Intervention[D]	261	Mexico	Supplements	*IGF2*	CBMCs	Hyper	9
GENOME-WIDE									
PUFA n-3	Amarasekera, 2014	Intervention[D]	70	Australia	Intervention (Fish oil)[#]	NA	Cord blood	No	8
CHILDHOOD									
GENOME-WIDE									
Fat	Voisin. 2014	Cross-sectional	69	Greece	24h recalls (3 d)	NA	Peripheral whole blood	Hyper/Hypo	5
MUFA/SFA	Voisin. 2014	Cross-sectional	69	Greece	24h recalls (3 d)	NA	Peripheral whole blood	Hyper/Hypo	5
PUFA/SFA	Voisin. 2014	Cross-sectional	69	Greece	24h recalls (3 d)	NA	Peripheral whole blood	Hyper/Hypo	5
(MUFA+PUFA)/SFA	Voisin. 2014	Cross-sectional	69	Greece	24h recalls (3 d)	NA	Peripheral whole blood	Hyper/Hypo	5
PUFA n-3	Lind, 2015	Intervention[D]	12	Denmark	Intervention (Fish oil)	NA	Buffy coats	No	8
ADULTHOOD									

Nutritional exposure	First author, year	Study design	N	Country	Exposure assessment	Genes	Tissue	Direction results	QS
GLOBAL									
Fat	Zhang 2012	Cross-sectional	165	USA	FFQ	NA	WBCs	No	4
Fat	Zhang, 2011	Cross-sectional	149	USA	FFQ	NA	Peripheral blood leukocytes	No	4
Fat	Ulrich, 2012	Cross-sectional	173	USA	FFQ	NA	Lymphocytes	Hypo	4
Fat (Lipids)	Gomes, 2012	Cross-sectional	126	Brazil	24h recall o(3 d)	NA	Venous blood	Hypo	3
MUFA	Zhang 2012	Cross-sectional	165	USA	FFQ	NA	WBCs	No	4
MUFA	Zhang, 2011	Cross-sectional	149	USA	FFQ	NA	Peripheral blood leukocytes	No	4
PUFA	Zhang 2012	Cross-sectional	165	USA	FFQ	NA	WBCs	No	4
PUFA	Zhang, 2011	Cross-sectional	149	USA	FFQ	NA	Peripheral blood leukocytes	No	4
SFA	Zhang 2012	Cross-sectional	165	USA	FFQ	NA	WBCs	No	4
SFA	Zhang, 2011	Cross-sectional	149	USA	FFQ	NA	Peripheral blood leukocytes	Hypo	4
GENE-SPECIFIC									
Fat (Animal)	Stidley, 2010	Cross-sectional	1101	USA	FFQ	Index**	Sputum	No	6
Fat	Zhang 2012	Cross-sectional	165	USA	FFQ	*IL-6*	WBCs	No	5
Fatty acids (circulating ALA)	Ma, 2016 - AJCN	Cross-sectional	991	USA	RBC	*APOE*	$CD4^+$ T cells (from buffy coats)	NA	7
Fatty acids (circulating EPA)	Ma, 2016 - AJCN	Cross-sectional	991	USA	RBC	*ABCA1*	$CD4^+$ T cells (from buffy coats)	NA	7
Fat (Lipids)	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*CD14*	Blood	No	6
Fat (Lipids)	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*Et-1*	Blood	No	6
Fat (Lipids)	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*HERV-w*	Blood	No	6
Fat (Lipids)	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*iNOS*	Blood	No	6
Fat (Lipids)	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*TNFα*	Blood	No	6
Fat (Lipids)	Gómez-Uriz, 2014	Cross-sectional	12	Spain	FFQ	*TNF-α*	Whole blood	Hypo	4
MUFA	Zhang 2012	Cross-sectional	165	USA	FFQ	*IL-6*	WBCs	No	5
MUFA	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*CD14*	Blood	No	6
MUFA	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*Et-1*	Blood	No	6
MUFA	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*HERV-w*	Blood	No	6

Nutritional exposure	First author, year	Study design	N	Country	Exposure assessment	Genes	Tissue	Direction results	QS
MUFA	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*iNOS*	Blood	No	6
MUFA	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*TNFα*	Blood	No	6
MUFA	Milagro, 2012	Cross-sectional	60	Spain	24h recall	*CLOCK*	WBCs	Hypo	4
PUFA	Zhang 2012	Cross-sectional	165	USA	FFQ	*IL-6*	WBCs	No	5
PUFA	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*CD14*	Blood	No	6
PUFA	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*Et-1*	Blood	No	6
PUFA	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*HERV-w*	Blood	No	6
PUFA	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*iNOS*	Blood	No	6
PUFA	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*TNFα*	Blood	No	6
PUFA	Milagro, 2012	Cross-sectional	60	Spain	24h recall	*CLOCK*	WBCs	Hyper	4
PUFA n-3	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*CD14*	Blood	No	6
PUFA n-3	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*Et-1*	Blood	No	6
PUFA n-3	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*HERV-w*	Blood	No	6
PUFA n-3	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*iNOS*	Blood	No	6
PUFA n-3	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*TNFα*	Blood	No	6
PUFA n-3	Ma, 2016 - MNFR	Cross-sectional	848	USA	RBC	*IL6*	$CD4^+$ T cells (from buffy coats)	Hypo	7
PUFA n-6	Hermsdorff, 2013	Cross-sectional	40	Spain	FFQ	*TNFα*	Blood	Hypo	3
Fat (Saturated)	Zhang 2012	Cross-sectional	165	USA	FFQ	*IL-6*	WBCs	No	5
SFA	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*CD14*	Blood	No	6
SFA	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*Et-1*	Blood	No	6
SFA	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*HERV-w*	Blood	No	6
SFA	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*iNOS*	Blood	No	6
SFA	Bollati, 2014	Cross-sectional	165	Italy	FFQ	*TNFα*	Blood	No	6
Fat	Stidley, 2010	Cross-sectional	1101	USA	FFQ	Index**	Sputum	No	6
Fat	Brøns, 2010	Intervention[N]	46	Denmark	Diet[##]	*PPARGC1A*	Muscle biopsie	Hypo	6
Fat	Gillberg, 2014	Intervention[N]	45	Denmark	Diet[###]	*PPARGC1A*	Subcutaneous adipose tissue	Hyper	6

Nutritional exposure	First author, year	Study design	N	Country	Exposure assessment	Genes	Tissue	Direction results	QS
GENOME-WIDE									
PUFA n-3	Aslibekyan, 2014	Cross-sectional	185	USA	RBCs	NA	Blood	Hyper/Hypo	6
Fat	Jacobsen, 2012	Intervention[N]	21	Denmark	Diet (Restricted high fat diet for 5 d)	NA	Skeletal Muscle biopsies	Hyper/Hypo	6
Fat	Jacobsen, 2014	Intervention[N]	40	Denmark	Diet (Restricted high fat diet for 5 d)	NA	Skeletal Muscle biopsies	Hyper/Hypo	6
Fat	Gillberg, 2016	Intervention[N]	34	Denmark	Diet[###] + 7 wk of overfeeding	NA	Subcutaneous adipose tissue	Hyper/Hypo	6
Fatty acids	Irvin, 2014	Intervention[N]	1048	USA	Diet (Restricted)	NA	Blood	No	8

*PUFA n-3 [docosahexaenoic acid (DHA)]

**Index *(p16, MGMT, DAPK, RASSF1A, PAX5α, PAX5β, GATA4, and GATA5)*

[#]Intervention; 3.7 g of fish oil (with 56.0% as DHA & 27.7% as EPA) or placebo in capsules daily from 20 weeks of gestation until delivery

[##]Diet (3d control diet including 30% fat & 5d high-fat overfeeding diet containing 50%) extra calories & 60% fat)

[###]Diet (5-d high-fat overfeeding diet (60E% fat, 50% extra calories) & a weight maintaining control diet (30E% fat); 6-8 wk of wash-out)

CBMCs (Cord blood mononuclear cells)

Table 17. Study characteristics and results: Carbohydrates and fiber

Nutritional exposure	First author, year	Study design	N	Country	Exposure assessment	Genes	Tissue	Direction results	QS
MATERNAL OFFSPRING RELATIONSHIP									
GLOBAL									
Carbohydrates	Rerkasem, 2015	Cross-sectional	249	Thailand	FFQ & 24h recall	NA	Venous blood	No	3
GENE-SPECIFIC									
Carbohydrates	Godfrey, 2011	Cross-sectional	78	UK	FFQ	RXRA	Cord tissue	Hypo	3
Carbohydrates	Godfrey, 2011	Cross-sectional	78	UK	FFQ	eNOS	Cord tissue	No	3
ADULTHOOD									
GLOBAL									
Carbohydrates	Gomes, 2012	Cross-sectional	126	Brazil	24h recall (3 d)	NA	Venous blood	Hypo	3
Carbohydrates	Zhang, 2011	Cross-sectional	149	USA	FFQ	NA	Peripheral blood leukocytes	No	4
Carbohydrates	Zhang 2012	Cross-sectional	165	USA	FFQ	NA	WBCs	No	4
Carbohydrates	Ulrich, 2012	Cross-sectional	173	USA	FFQ	NA	Lymphocytes	No	4
GENE-SPECIFIC									
Carbohydrates	Samblas, 2016	Cross-sectional	61	Spain	Questionnaire	BMAL1	Whole blood	Hyper	4
Carbohydrates	Bollati, 2014	Cross-sectional	165	Italy	FFQ	CD14	Blood	No	6
Carbohydrates	Bollati, 2014	Cross-sectional	165	Italy	FFQ	Et-1	Blood	No	6
Carbohydrates	Bollati, 2014	Cross-sectional	165	Italy	FFQ	HERV-w	Blood	No	6
Carbohydrates	Bollati, 2014	Cross-sectional	165	Italy	FFQ	iNOS	Blood	No	6
Carbohydrates	Bollati, 2014	Cross-sectional	165	Italy	FFQ	TNFα	Blood	No	6
Carbohydrates	Zhang 2012	Cross-sectional	165	USA	FFQ	IL-6	WBCs	No	5
Fiber	Bollati, 2014	Cross-sectional	165	Italy	FFQ	CD14	Blood	No	6
Fiber	Bollati, 2014	Cross-sectional	165	Italy	FFQ	Et-1	Blood	No	6
Fiber	Bollati, 2014	Cross-sectional	165	Italy	FFQ	HERV-w	Blood	No	6
Fiber	Bollati, 2014	Cross-sectional	165	Italy	FFQ	iNOS	Blood	No	6
Fiber	Bollati, 2014	Cross-sectional	165	Italy	FFQ	TNFα	Blood	No	6

Table 18. Study characteristics and results: Protein and amino acids

Nutritional exposure	First author, year	Study design	N	Country	Exposure assessment	Genes	Tissue	Direction results	QS
MATERNAL OFFSPRING RELATIONSHIP									
GLOBAL									
Protein	Rerkasem, 2015	Cross-sectional	249	Thailand	FFQ & 24h recall	NA	Venous blood samples in children	No	3
GENE-SPECIFIC									
Protein	Godfrey, 2011	Cross-sectional	78	UK	FFQ	RXRA	Cord tissue	No	3
Protein	Godfrey, 2011	Cross-sectional	78	UK	FFQ	eNOS	Cord tissue	No	3
ADULTHOOD									
GLOBAL									
Protein	Gomes, 2012	Cross-sectional	126	Brazil	24h recall (3 d)	NA	Venous blood	No	3
Protein	Zhang 2012	Cross-sectional	165	USA	FFQ	NA	WBCs	No	4
Protein	Zhang, 2011	Cross-sectional	149	USA	FFQ	NA	Peripheral blood leukocytes	No	4
Methionine	Huang, 2012	Cross-sectional	493	USA	FFQ	NA	Blood	No	3
Methionine	Perng, 2014	Cross-sectional	987	USA	FFQ	NA	Blood sample	No	3
Methionine	Zhang 2012	Cross-sectional	165	USA	FFQ	NA	WBCs	No	4
Methionine	Zhang, 2011	Cross-sectional	149	USA	FFQ	NA	Peripheral blood leukocytes	No	4
GENE-SPECIFIC									
Protein	Bollati, 2014	Cross-sectional	165	Italy	FFQ	CD14	Blood	No	6
Protein	Bollati, 2014	Cross-sectional	165	Italy	FFQ	Et-1	Blood	No	6
Protein	Bollati, 2014	Cross-sectional	165	Italy	FFQ	HERV-w	Blood	No	6
Protein	Bollati, 2014	Cross-sectional	165	Italy	FFQ	iNOS	Blood	No	6
Protein	Bollati, 2014	Cross-sectional	165	Italy	FFQ	TNFα	Blood	No	6
Protein	Zhang 2012	Cross-sectional	165	USA	FFQ	IL-6	WBCs	No	5
Methionine	Zhang 2012	Cross-sectional	165	USA	FFQ	IL-6	WBCs	No	5
Methionine	Zhong, 2015	Cross-sectional	573	USA	FFQ	TLR2	Whole blood	No	6

SUPPLEMENTARY MATERIAL

Supplement 1: Details on the search strategy for each of the databases

	References identified	Unique references identified
Embase.com	2337	2294
Medline Ovid	1547	249
Web of science	1714	943
Cochrane	87	19
PubMed publisher	149	133
Google scholar	200	136
Total	**5874**	**3774**

Embase.com 2337

(diet/exp OR 'dietary intake'/exp OR 'food intake'/de OR 'caloric density'/exp OR 'caloric intake'/exp OR 'carbohydrate intake'/exp OR 'diet restriction'/exp OR 'dietary reference intake'/exp OR 'electrolyte intake'/exp OR 'fat intake'/exp OR 'fluid intake'/exp OR 'food deprivation'/exp OR 'mineral intake'/exp OR 'protein intake'/exp OR 'vitamin intake'/exp OR 'vitamin blood level'/exp OR 'diet supplementation'/exp OR 'mineral supplementation'/exp OR 'vitamin supplementation'/exp OR supplementation/de OR 'nutrition'/de OR 'child nutrition'/de OR 'diet therapy'/exp OR 'food intake'/exp OR 'infant nutrition'/exp OR 'maternal nutrition'/de OR nutrient/exp OR 'nutritional status'/exp OR (diet OR nutrition OR ((dietar* OR food* OR beverage* OR alcohol* OR mineral* OR nutrient* OR micronutrient* OR macronutrient* OR vitamin* OR calor* OR energ* OR protein* OR fat OR folic-acid* OR folate* OR vegetabl* OR fruit* OR cholin* OR methionin* OR betain* OR fatty-acid* OR pufa OR pufas OR mufa OR mufas OR safa OR safas OR sfa OR sfas) NEAR/6 (intake* OR ingest* OR supplement* OR consum* OR restrict* OR depriv* OR level OR concentration OR blood OR plasma OR serum OR pattern*)) OR fasting OR ((well OR under) NEXT/1 nourish*) OR ((well OR under) NEXT/1 nourish*)):ab,ti) AND ('DNA methylation'/exp OR (((dna OR 'long interspersed' OR gene OR genes) NEAR/6 (demethylat* OR methylat* OR hypermethylat* OR hypomethylat*))):ab,ti) NOT ([animals]/lim NOT [humans]/lim) AND ('observational study'/exp OR 'cohort analysis'/exp OR 'longitudinal study'/exp OR 'retrospective study'/exp OR 'prospective study'/exp OR 'health survey'/de OR 'health care survey'/de OR 'epidemiological data'/de OR 'case control study'/de OR 'cross-sectional study'/de OR 'correlational study'/de OR 'population research'/de OR 'family study'/de OR 'major clinical study'/de OR 'multicenter study'/de OR 'comparative study'/de OR 'follow up'/de OR 'clinical study'/de OR 'clinical article'/de OR 'clinical trial'/exp OR 'randomization'/exp OR 'intervention study'/de OR 'open study'/de OR 'community trial'/de OR 'review'/exp OR 'systematic review'/exp OR (((observation* OR epidemiolog* OR famil* OR comparativ* OR communit*) NEAR/6 (stud* OR data OR research)) OR cohort* OR longitudinal* OR retrospectiv* OR prospectiv* OR population* OR (national* NEAR/3 (stud* OR survey))

OR (health* NEAR/3 survey*) OR ((case OR cases OR match*) NEAR/3 control*) OR (cross NEXT/1 section*) OR correlation* OR multicenter* OR multi-center* OR follow-up* OR followup* OR clinical* OR trial OR random* OR review* OR meta-analy*):ab,ti) NOT ([Conference Abstract]/lim OR [Letter]/lim OR [Note]/lim OR [Editorial]/lim)

Medline Ovid 1547

(exp Diet/ OR exp "Eating"/ OR "Recommended Dietary Allowances"/ OR "food deprivation"/ OR vitamins/bl OR "Dietary Supplements"/ OR "Nutritional Physiological Phenomena"/ OR exp "Diet Therapy"/ OR "Maternal Nutritional Physiological Phenomena"/ OR "nutritional status"/ OR (diet OR nutrition OR ((dietar* OR food* OR beverage* OR alcohol* OR mineral* OR nutrient* OR micronutrient* OR macronutrient* OR vitamin* OR calor* OR energ* OR protein* OR fat OR folic-acid* OR folate* OR vegetabl* OR fruit* OR cholin* OR methionin* OR betain* OR fatty-acid* OR pufa OR pufas OR mufa OR mufas OR safa OR safas OR sfa OR sfas) ADJ6 (intake* OR ingest* OR supplement* OR consum* OR restrict* OR depriv* OR level OR concentration OR blood OR plasma OR serum OR pattern*)) OR fasting OR ((well OR under) ADJ nourish*) OR ((well OR under) ADJ nourish*)).ab,ti.) AND ("DNA Methylation"/ OR (((dna OR "long interspersed" OR gene OR genes) ADJ6 (demethylat* OR methylat* OR hypermethylat* OR hypomethylat*))).ab,ti.) NOT (exp animals/ NOT humans/) AND ("observational study"/ OR exp "Epidemiologic Studies"/ OR "health surveys"/ OR "multicenter study"/ OR exp "clinical study"/ OR "Random Allocation"/ OR "review"/ OR (((observation* OR epidemiolog* OR famil* OR comparativ* OR communit*) ADJ6 (stud* OR data OR research)) OR cohort* OR longitudinal* OR retrospectiv* OR prospectiv* OR population* OR (national* ADJ3 (stud* OR survey)) OR (health* ADJ3 survey*) OR ((case OR cases OR match*) ADJ3 control*) OR (cross ADJ section*) OR correlation* OR multicenter* OR multi-center* OR follow-up* OR followup* OR clinical* OR trial OR random* OR review* OR meta-analy*).ab,ti.) NOT (letter OR news OR comment OR editorial OR congresses OR abstracts).pt.

Cochrane 87

((diet OR nutrition OR ((dietar* OR food* OR beverage* OR alcohol* OR mineral* OR nutrient* OR micronutrient* OR macronutrient* OR vitamin* OR calor* OR energ* OR protein* OR fat OR folic-acid* OR folate* OR vegetabl* OR fruit* OR cholin* OR methionin* OR betain* OR fatty-acid* OR pufa OR pufas OR mufa OR mufas OR safa OR safas OR sfa OR sfas) NEAR/6 (intake* OR ingest* OR supplement* OR consum* OR restrict* OR depriv* OR level OR concentration OR blood OR plasma OR serum OR pattern*)) OR fasting OR ((well OR under) NEXT/1 nourish*) OR ((well OR under) NEXT/1 nourish*)):ab,ti) AND ((((dna OR 'long interspersed' OR gene OR genes) NEAR/6 (demethylat* OR methylat* OR hypermethylat* OR hypomethylat*))):ab,ti)

Web of science 1714

TS=(((diet OR nutrition OR ((dietar* OR food* OR beverage* OR alcohol* OR mineral* OR nutrient* OR micronutrient* OR macronutrient* OR vitamin* OR calor* OR energ* OR protein* OR fat OR folic-acid* OR folate* OR vegetabl* OR fruit* OR cholin* OR methionin* OR betain* OR fatty-acid* OR pufa OR pufas OR mufa OR mufas OR safa OR safas OR sfa OR sfas) NEAR/5 (intake* OR ingest* OR supplement* OR consum* OR restrict* OR depriv* OR level OR concentration OR blood OR plasma OR serum OR pattern*)) OR fasting OR ((well OR under) NEAR/1 nourish*) OR ((well OR under) NEAR/1 nourish*))) AND ((((dna OR "long interspersed" OR gene OR genes) NEAR/5 (demethylat* OR methylat* OR hypermethylat* OR hypomethylat*)))) AND ((((observation* OR epidemiolog* OR famil* OR comparativ* OR communit*) NEAR/5 (stud* OR data OR research)) OR cohort* OR longitudinal* OR retrospectiv* OR prospectiv* OR population* OR (national* NEAR/2 (stud* OR survey)) OR (health* NEAR/2 survey*) OR ((case OR cases OR match*) NEAR/2 control*) OR (cross NEAR/1 section*) OR correlation* OR multicenter* OR multi-center* OR follow-up* OR followup* OR clinical* OR trial OR random* OR review* OR meta-analy*)) NOT ((animal* OR rat OR rats OR mouse OR mice OR murine) NOT (human* OR patient*))) AND DT=(article)

PubMed publisher 149

(Diet[mh] OR "Eating"[mh] OR "Recommended Dietary Allowances"[mh] OR "food deprivation"[mh] OR vitamins/bl[mh] OR "Dietary Supplements"[mh] OR "Nutritional Physiological Phenomena"[mh] OR "Diet Therapy"[mh] OR "Maternal Nutritional Physiological Phenomena"[mh] OR "nutritional status"[mh] OR (diet OR nutrition OR ((dietar*[tiab] OR food*[tiab] OR beverage*[tiab] OR alcohol*[tiab] OR mineral*[tiab] OR nutrient*[tiab] OR micronutrient*[tiab] OR macronutrient*[tiab] OR vitamin*[tiab] OR calor*[tiab] OR energ*[tiab] OR protein*[tiab] OR fat OR folic-acid*[tiab] OR folate*[tiab] OR vegetabl*[tiab] OR fruit*[tiab] OR cholin*[tiab] OR methionin*[tiab] OR betain*[tiab] OR fatty-acid*[tiab] OR pufa OR pufas OR mufa OR mufas OR safa OR safas OR sfa OR sfas) AND (intake*[tiab] OR ingest*[tiab] OR supplement*[tiab] OR consum*[tiab] OR restrict*[tiab] OR depriv*[tiab] OR level OR concentration OR blood OR plasma OR serum OR pattern*[tiab])) OR fasting OR well nourish*[tiab] OR under nourish*[tiab])) AND ("DNA Methylation"[mh] OR (((dna OR "long interspersed" OR gene OR genes) AND (demethylat*[tiab] OR methylat*[tiab] OR hypermethylat*[tiab] OR hypomethylat*[tiab])))) NOT (animals[mh] NOT humans[mh]) AND ("observational study"[mh] OR "Epidemiologic Studies"[mh] OR "health surveys"[mh] OR "multicenter study"[mh] OR "clinical study"[mh] OR "Random Allocation"[mh] OR "review"[mh] OR (((observation*[tiab] OR epidemiolog*[tiab] OR famil*[tiab] OR comparativ*[tiab] OR communit*[tiab]) AND (stud*[tiab] OR data OR research)) OR cohort*[tiab] OR longitudinal*[tiab] OR retrospectiv*[tiab] OR prospectiv*[tiab] OR population*[tiab]

OR (national*[tiab] AND (stud*[tiab] OR survey)) OR (health*[tiab] AND survey*[tiab]) OR ((case OR cases OR match*[tiab]) AND control*[tiab]) OR (cross section*[tiab]) OR correlation*[tiab] OR multicenter*[tiab] OR multi-center*[tiab] OR follow-up*[tiab] OR followup*[tiab] OR clinical*[tiab] OR trial OR random*[tiab] OR review*[tiab] OR meta-analy*[tiab])) NOT (letter[pt] OR news[pt] OR comment[pt] OR editorial[pt] OR congresses[pt] OR abstracts[pt]) AND publisher[sb]

Google scholar

diet|nutrition|dietary|nutrients|micronutrients|macronutrients|vitamins|"folic acid"|folate|vegetables|fruit|choline
"dna demethylation|methylation|hypermethylation|hypomethylation"

Supplement 2: Detailed selection criteria

Inclusion criteria	1. Any study design • Including cross-sectional studies, case-control studies, cohort studies or intervention studies. 2. Studies investigating associations of dietary factors with DNA methylation. • Nutrients: Including nutrient blood levels; intake of nutrients; or dietary supplements. • DNA methylation: Including global DNA methylation, gene specific DNA methylation, or genome-wide DNA methylation. 3. Studies conducted at any age or life stage (e.g., during pregnancy, childhood, adulthood, etc.) 4. No language or date restriction
Exclusion criteria	1. Studies not carried out in humans 2. In vitro studies 3. Studies conducted in only patients with chronic diseases (cancer, diabetes, CHD, etc.). Note: some studies have healthy control groups and report associations within these groups separately, these should be included. 4. Letters, reviews, opinions papers, guidelines, case-reports (n<5), editorials 5. Studies examining only the following exposures: • eating disorders; • alcohol dependency/alcoholism (but not actual alcohol intake); • weight loss (unless the weight-loss intervention is solely a restricted diet, not combined with other lifestyle changes). • Dietary patterns or food groups without examining individual nutrients

Supplement 3: Quality score

Two independent reviewers evaluated the quality of included studies using the following quality score. The score is composed of 5 items, and each item was allocated 0, 1 or 2 points. This allowed a total score between 0 and 10 points, 10 representing the highest quality. The following items are included in the score:

1. **Study design**
   - **0** for cross-sectional studies
   - **1** for longitudinal studies (including repeated measurements)
   - **2** for interventional studies
2. **Population**
   - **0** if *n*<30
   - **1** if *n* 30 to 100
   - **2** if *n*>100
3. **Exposure**
   - ***Observational studies***
     - **0** if the study used no appropriate standard diet or nutrient concentration assessment method (see below) or if not reported
     - **1** if the study used food records, a 24h recall or an FFQ
     - **2** if the study used blood levels
   - ***Interventional studies***
     - **0** if the intervention diet was not described or not blinded
     - **1** if the intervention diet was adequately single-blinded
     - **2** if the intervention is adequately double-blinded
4. **Outcome**
   - **0** Global DNA methylation
   - **1** Gene-specific DNA methylation
   - **2** Genome-wide DNA methylation
5. **Adjustments**
   - **0** if findings are not controlled for potential confounders*
   - **1** if findings are controlled for at least basic confounders: age and sex, and multiple testing in case of an EWAS
   - **2** if an intervention study is adequately randomized or if findings are controlled for additional confounders (on top of those described above), such as: technical covariates, WBC, BMI, smoking, alcohol, etc.

** 'Controlled for' includes: adjustment for in the statistical analyses (e.g. with multivariable regression); stratification in the analyses (e.g. men and women separately); or restriction or narrow selection criteria of study participants on this covariate.*

REFERENCES

1. Richmond RC, Timpson NJ, Sorensen TI. Exploring possible epigenetic mediation of early-life environmental exposures on adiposity and obesity development. Int J Epidemiol. 2015;44(4):1191-8.
2. Friso S, Udali S, De Santis D, Choi SW. One-carbon metabolism and epigenetics. Mol Aspects Med. 2017;54:28-36.
3. Milagro FI, Mansego ML, De Miguel C, Martínez JA. Dietary factors, epigenetic modifications and obesity outcomes: Progresses and perspectives. Mol Asp Med. 2013;34(4):782-812.
4. Anderson OS, Sant KE, Dolinoy DC. Nutrition and epigenetics: An interplay of dietary methyl donors, one-carbon metabolism and DNA methylation. J Nutr Biochem. 2012;23(8):853-9.
5. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. Br Med J 2009;339:b2700.
6. Voortman T, Vitezova A, Bramer WM, Ars CL, Bautista PK, Buitrago-Lopez A, et al. Effects of protein intake on blood pressure, insulin sensitivity, and blood lipids in children: a systematic review. Br J Nutr. 2015;113(3):In press.
7. Perng W, Rozek LS, Mora-Plazas M, Duchin O, Marin C, Forero Y, et al. Micronutrient status and global DNA methylation in school-age children. Epigenetics. 2012;7(10):1133-41.
8. Piyathilake CJ, Macaluso M, Alvarez RD, Chen M, Badiga S, Siddiqui NR, et al. A higher degree of LINE-1 methylation in peripheral blood mononuclear cells, a one-carbon nutrient related epigenetic alteration, is associated with a lower risk of developing cervical intraepithelial neoplasia. Nutrition. 2011;27(5):513-9.
9. Bollati V, Favero C, Albetti B, Tarantini L, Moroni A, Byun HM, et al. Nutrients intake is associated with DNA methylation of candidate inflammatory genes in a population of obese subjects. Nutrients. 2014;6(10):4625-39.
10. Stidley CA, Picchi MA, Leng S, Willink R, Crowell RE, Flores KG, et al. Multivitamins, folate, and green vegetables protect against gene promoter methylation in the aerodigestive tract of smokers. Cancer Res. 2010;70(2):568-74.
11. Haggarty P, Hoad G, Campbell DM, Horgan GW, Piyathilake C, McNeill G. Folate in pregnancy and imprinted gene and repeat element methylation in the offspring. Am J Clin Nutr. 2013;97(1):94-9.
12. Boeke CE, Baccarelli A, Kleinman KP, Burris HH, Litonjua AA, Rifas-Shiman SL, et al. Gestational intake of methyl donors and global LINE-1 DNA methylation in maternal and cord blood: Prospective results from a folate-replete population. Epigenetics. 2012;7(3):253-60.
13. Fryer AA, Nafee TM, Ismail KM, Carroll WD, Emes RD, Farrell WE. LINE-1 DNA methylation is inversely correlated with cord plasma homocysteine in man: a preliminary study. Epigenetics. 2009;4(6):394-8.
14. McKay JA, Groom A, Potter C, Coneyworth LJ, Ford D. Genetic and non-genetic influences during pregnancy on infant global and site specific DNA methylation: role for folate gene variants and vitamin B 12: journals.plos.org; 2012.
15. Chang H, Zhang T, Zhang Z, Bao R, Fu C. Tissue-specific distribution of aberrant DNA methylation associated with maternal low-folate status in human neural tube defects. The Journal of nutritional …. 2011.
16. Park BH, Kim YJ, Park JS, Lee HY, Ha EH, Min JW, et al. Folate and homocysteine levels during pregnancy affect DNA methylation in human placenta. J Prev Med Pub Health. 2005;38(4):437-42.

17. Ba Y, Yu H, Liu F, Geng X, Zhu C, Zhu Q, et al. Relationship of folate, vitamin B12 and methylation of insulin-like growth factor-II in maternal and cord blood. Eur J Clin Nutr. 2011;65(4):480-5.
18. Hoyo C, Daltveit AK, Iversen E, Benjamin-Neelon SE, Fuemmeler B, Schildkraut J, et al. Erythrocyte folate concentrations, CpG methylation at genomically imprinted domains, and birth weight in a multiethnic newborn cohort. Epigenetics. 2014;9(8):1120-30.
19. Steegers-Theunissen RP, Obermann-Borst SA. Periconceptional maternal folic acid use of 400 µg per day is related to increased methylation of the IGF2 gene in the very young child: journals.plos.org; 2009.
20. Hoyo C, Murtha AP, Schildkraut JM, Jirtle R, Demark-Wahnefried W, Forman MR, et al. Methylation variation at IGF2 differentially methylated regions and maternal folic acid use before and during pregnancy. Epigenetics. 2011;6(7):928-36.
21. Van Mil NH, Bouwl-Both MI, Stolk L, Verbiest MMPJ, Hofman A, Jaddoe VWV, et al. Determinants of maternal pregnancy one-carbon metabolism and newborn human DNA methylation profiles. Reproduction. 2014;148(6):581-92.
22. Azzi S, Sas TCJ, Koudou Y, Le Bouc Y, Souberbielle JC, Dargent-Molina P, et al. Degree of methylation of ZAC1 (PLAGL1) is associated with prenatal and post-natal growth in healthy infants of the EDEN mother child cohort. Epigenetics. 2014;9(3):338-45.
23. Amarasekera M, Martino D, Ashley S, Harb H, Kesper D, Strickland D, et al. Genome-wide DNA methylation profiling identifies a folate-sensitive region of differential methylation upstream of ZFP57-imprinting regulator in humans. FASEB J. 2014;28(9):4068-76.
24. Joubert BR, Den Dekker HT, Felix JF, Bohlin J, Ligthart S, Beckett E, et al. Maternal plasma folate impacts differential DNA methylation in an epigenome-wide meta-analysis of newborns. Nat Commun. 2016;7.
25. Gonseth S, Roy R, Houseman EA, de Smith AJ, Zhou M, Lee ST, et al. Periconceptional folate consumption is associated with neonatal DNA methylation modifications in neural crest regulatory and cancer development genes. Epigenetics. 2015;10(12):1166-76.
26. Mozhui K, Smith AK, Tylavsky FA. Ancestry Dependent DNA Methylation and Influence of Maternal Nutrition. PLoS ONE. 2015;10(3).
27. Ono H, Iwasaki M, Kuchiba A, Kasuga Y, Yokoyama S, Onuma H, et al. Association of dietary and genetic factors related to one-carbon metabolism with global methylation level of leukocyte DNA. Cancer Sci. 2012;103(12):2159-64.
28. Agodi A, Barchitta M, Quattrocchi A, Maugeri A, Canto C, Marchese AE, et al. Low fruit consumption and folate deficiency are associated with LINE-1 hypomethylation in women of a cancer-free population. Genes Nutr. 2015;10(5).
29. Huang WY, Su LJ, Hayes RB, Moore LE, Katki HA, Berndt SI, et al. Prospective study of genomic hypomethylation of leukocyte DNA and colorectal cancer risk. Cancer Epidemiol Biomarkers Prev. 2012;21(11):2014-21.
30. Zhang FF, Santella RM, Wolff M, Kappil MA, Markowitz SB, Morabia A. White blood cell global methylation and IL-6 promoter methylation in association with diet and lifestyle risk factors in a cancer-free population. Epigenetics. 2012;7(6):606-14.
31. Gomes MVM, Toffoli LV, Arruda DW, Soldera LM, Pelosi GG, Neves-Souza RD, et al. Age-Related Changes in the Global DNA Methylation Profile of Leukocytes Are Linked to Nutrition but Are Not Associated with the MTHFR C677T Genotype or to Functional Capacities. PLoS ONE. 2012;7(12).
32. Perng W, Villamor E, Shroff MR, Nettleton JA, Pilsner JR, Liu Y, et al. Dietary intake, plasma homocysteine, and repetitive element DNA methylation in the Multi-Ethnic Study of Atherosclerosis (MESA). Nutr Metab Cardiovasc Dis. 2014;24(6):614-22.

33. Protiva P, Mason JB, Liu Z, Hopkins ME, Nelson C, Marshall JR, et al. Altered folate availability modifies the molecular environment of the human colorectum: Implications for colorectal carcinogenesis. Cancer Prev Res. 2011;4(4):530-43.
34. Ulrich CM, Toriola AT, Koepl LM, Sandifer T, Poole EM, Duggan C, et al. Metabolic, hormonal and immunological associations with global DNA methylation among postmenopausal women. Epigenetics. 2012;7(9):1020-8.
35. Zhang FF, Morabia A, Carroll J, Gonzalez K, Fulda K, Kaur M, et al. Dietary patterns are associated with levels of global genomic DNA methylation in a cancer-free population. J Nutr. 2011;141(6):1165-71.
36. Badiga S, Johanning GL, Macaluso M, Azuero A, Chambers MM, Siddiqui NR, et al. A lower degree of PBMC L1 methylation in women with lower folate status may explain the MTHFR C677T polymorphism associated higher risk of CIN in the US post folic acid fortification era. PLoS ONE. 2014;9(10).
37. Bae S, Ulrich CM, Bailey LB, Malysheva O, Brown EC, Maneval DR, et al. Impact of folic acid fortification on global DNA methylation and one-carbon biomarkers in the Women's Health Initiative Observational Study cohort. Epigenetics. 2014;9(3):396-403.
38. Friso S, Choi SW, Girelli D, Mason JB. A common mutation in the 5, 10-methylenetetrahydrofolate reductase gene affects genomic DNA methylation through an interaction with folate status. Proceedings of the …. 2002.
39. Friso S, Girelli D, Trabetti E, Olivieri O, Guarini P, Pignatti PF, et al. The MTHFR 1298A>C polymorphism and genomic DNA methylation in human lymphocytes. Cancer Epidemiol Biomarkers Prev. 2005;14(4):938-43.
40. Pilsner JR, Liu X, Ahsan H, Ilievski V, Slavkovich V, Levy D, et al. Genomic methylation of peripheral blood leukocyte DNA: Influences of arsenic and folate in Bangladeshi adults. Am J Clin Nutr. 2007;86(4):1179-86.
41. Piyathilake CJ, Badiga S, Alvarez RD, Partridge EE, Johanning GL. A Lower Degree of PBMC L1 Methylation Is Associated with Excess Body Weight and Higher HOMA-IR in the Presence of Lower Concentrations of Plasma Folate. PLoS ONE. 2013;8(1).
42. Wang TC, Song YS, Wang H, Zhang J, Yu SF, Gu YE, et al. Oxidative DNA damage and global DNA hypomethylation are related to folate deficiency in chromate manufacturing workers. J Hazard Mater. 2012;213-214:440-6.
43. Fenech M, Aitken C, Rinaldi J. Folate, vitamin B12, homocysteine status and DNA damage in young Australian adults. Carcinogenesis. 1998;19(7):1163-71.
44. Gadgil MS, Joshi KS, Naik SS, Pandit AN, Otiv SR, Patwardhan BK. Association of homocysteine with global DNA methylation in vegetarian Indian pregnant women and neonatal birth anthropometrics. J Matern -Fetal Neonatal Med. 2014;27(17):1749-53.
45. Narayanan S, McConnell J, Little J, Sharp L, Piyathilake CJ, Powers H, et al. Associations between two common variants C677T and A1298C in the methylenetetrahydrofolate reductase gene and of measures of folate metabolism and DNA stability (strand breaks, misincorporated uracil, and DNA methylation status) in human lymphocytes in vivo. Cancer Epidemiol Biomarkers Prev. 2004;13(9):1436-43.
46. Axume J, Smith SS, Pogribny IP, Moriarty DJ, Caudill MA. The methylenetetrahydrofolate reductase 677TT genotype and folate intake interact to lower global leukocyte DNA methylation in young Mexican American women. Nutr Res. 2007;27(1):13-7.
47. Crider KS, Quinlivan EP, Berry RJ, Hao L, Li Z, Maneval D, et al. Genomic DNA methylation changes in response to folic acid supplementation in a population-based intervention study among women of reproductive age. PLoS ONE. 2011;6(12).
48. Jacob RA, Gretz DM, Taylor PC, James SJ. Moderate folate depletion increases plasma homocysteine and decreases lymphocyte DNA methylation in postmenopausal women. Journal of nutrition. 1998.

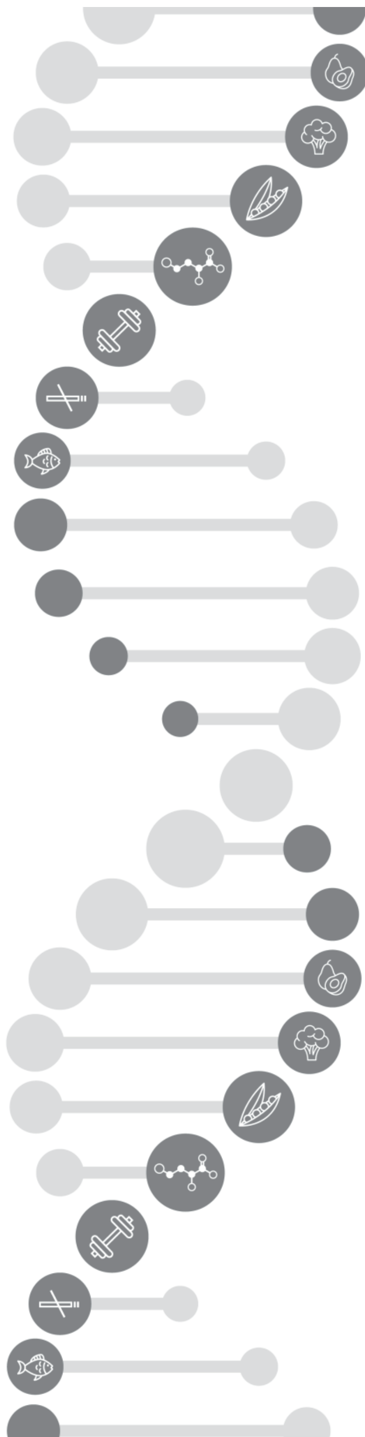
49. Shelnutt KP, Kauwell GPA, Gregory JF. Methylenetetrahydrofolate reductase 677C→ T polymorphism affects DNA methylation in response to controlled folate intake in young women. The Journal of nutritional …. 2004.
50. Rampersaud GC, Kauwell GPA, Hutson AD, Cerda JJ, Bailey LB. Genomic DNA methylation decreases in response to moderate folate depletion in elderly women. Am J Clin Nutr. 2000;72(4):998-1003.
51. Hubner U, Geisel J, Kirsch SH, Kruse V, Bodis M, Klein C, et al. Effect of 1 year B and D vitamin supplementation on LINE-1 repetitive element methylation in older subjects. Clin Chem Lab Med. 2013;51(3):649-55.
52. Jung AY, Smulders Y, Verhoef P, Kok FJ, Blom H, Kok RM, et al. No effect of folic acid supplementation on global DNA methylation in men and women with moderately elevated homocysteine. PLoS ONE. 2011;6(9).
53. Basten GP, Duthie SJ, Pirie L, Vaughan N, Hill MH, Powers HJ. Sensitivity of markers of DNA stability and DNA repair activity to folate supplementation in healthy volunteers. Br J Cancer. 2006;94(12):1942-7.
54. Figueiredo JC, Grau MV, Wallace K, Levine AJ, Shen L, Hamdan R, et al. Global DNA hypomethylation (LINE-1) in the normal colon and lifestyle characteristics and dietary and genetic factors. Cancer Epidemiol Biomarkers Prev. 2009;18(4):1041-9.
55. Hanks J, Ayed I, Kukreja N, Rogers C, Harris J, Gheorghiu A, et al. The association between mthfr 677C>T genotype and folate status and genomic and gene-Specific dna methylation in the colon of individuals without colorectal neoplasia1-4. Am J Clin Nutr. 2013;98(6):1564-74.
56. Pufulete M, Al-Ghnaniem R, Rennie JA, Appleby P, Harris N, Gout S, et al. Influence of folate status on genomic DNA methylation in colonic mucosa of subjects without colorectal adenoma or cancer. Br J Cancer. 2005;92(5):838-42.
57. Tapp HS, Commane DM, Bradburn DM, Arasaradnam R, Mathers JC, Johnson IT, et al. Nutritional factors and gender influence age-related DNA methylation in the human rectal mucosa. Aging Cell. 2013;12(1):148-55.
58. Llanos AAM, Marian C, Brasky TM, Dumitrescu RG, Liu Z, Mason JB, et al. Associations between genetic variation in onecarbon metabolism and LINE-1 DNA methylation in histologically normal breast tissues. Epigenetics. 2015;10(8):727-35.
59. Aarabi M, San Gabrie MC, Chan D, Behan NA, Caron M, Pastinen T, et al. High-dose folic acid supplementation alters the human sperm methylome and is influenced by the MTHFR C677T polymorphism. Hum Mol Genet. 2015;24(22):6301-13.
60. Zhong J, Colicino E, Lin X, Mehta A, Kloog I, Zanobetti A, et al. Cardiac autonomic dysfunction: Particulate air pollution effects are modulated by epigenetic immunoregulation of Toll-like receptor 2 and dietary flavonoid intake. J Am Heart Assoc. 2015;4(1).
61. Dhillon VS, Shahid M, Husain SA. Associations of MTHFR DNMT3b 4977 bp deletion in mtDNA and GSTM1 deletion, and aberrant CpG island hypermethylation of GSTM1 in non-obstructive infertility in Indian men. Mol Hum Reprod. 2007;13(4):213-22.
62. Ottini L, Rizzolo P, Siniscalchi E, Zijno A, Silvestri V, Crebelli R, et al. Gene promoter methylation and DNA repair capacity in monozygotic twins with discordant smoking habits. Mutat Res Genet Toxicol Environ Mutagen. 2015;779:57-64.
63. Piyathilake CJ, Macaluso M, Chambers MM, Badiga S, Siddiqui NR, Bell WC, et al. Folate and vitamin B12 may play a critical role in lowering the HPV 16 methylation - Associated risk of developing higher grades of CIN. Cancer Prev Res. 2014;7(11):1128-37.
64. Llanos AA, Dumitrescu RG, Brasky TM, Liu Z, Mason JB, Marian C, et al. Relationships among folate, alcohol consumption, gene variants in one-carbon metabolism and

p16INK4a methylation and expression in healthy breast tissues. Carcinogenesis. 2015;36(1):60-7.

65. Al-Ghnaniem Abbadi R, Emery P, Pufulete M. Short-Term Folate Supplementation in Physiological Doses Has No Effect on ESR1 and MLH1 Methylation in Colonic Mucosa of Individuals with Adenoma. J Nutrigenet Nutrigenomics. 2013;5(6):327-38.
66. Song MA, Brasky TM, Marian C, Weng DY, Taslim C, Llanos AA, et al. Genetic variation in one-carbon metabolism in relation to genome-wide DNA methylation in breast tissue from healthy women. Carcinogenesis. 2016;37(5):471-80.
67. Marques-Rocha JL, Milagro FI, Mansego ML, Mourão DM, Martínez JA, Bressan J. LINE-1 methylation is positively associated with healthier lifestyle but inversely related to body fat mass in healthy young individuals. Epigenetics. 2016;11(1):49-60.
68. Quinlivan EP, Crider KS, Zhu JH, Maneval DR, Hao L, Li Z, et al. Hypomethylation of Serum Blood Clot DNA, but Not Plasma EDTA-Blood Cell Pellet DNA, from Vitamin B12-Deficient Subjects. PLoS ONE. 2013;8(6).
69. Irvin MR, Zhi D, Aslibekyan S, Claas SA, Absher DM, Ordovas JM, et al. Genomics of post-prandial lipidomic phenotypes in the Genetics of Lipid Lowering Drugs and Diet Network (GOLDN) study. PLoS ONE. 2014;9(6).
70. Al-Ghnaniem R, Peters J, Foresti R, Heaton N, Pufulete M. Methylation of estrogen receptor α and mutL homolog 1 in normal colonic mucosa: Association with folate and vitamin B-12 status in subjects with and without colorectal neoplasia. Am J Clin Nutr. 2007;86(4):1064-72.
71. de la Iglesia R, Mansego ML, Sánchez-Muniz FJ, Angeles Zulet M, Alfredo Martinez J. Arylesterase activity is associated with antioxidant intake and Paraoxonase-1 (PON1) gene methylation in metabolic syndrome patients following an energy restricted diet. EXCLI J. 2014;13:416-26.
72. Harvey NC, Sheppard A, Godfrey KM, McLean C, Garratt E, Ntani G, et al. Childhood bone mineral content is associated with methylation status of the RXRA promoter at birth. J Bone Miner Res. 2014;29(3):600-7.
73. Novakovic B, Galati JC, Chen A, Morley R, Craig JM, Saffery R. Maternal vitamin D predominates over genetic factors in determining neonatal circulating vitamin D concentrations. Am J Clin Nutr. 2012;96(1):188-95.
74. Zhu H, Bhagatwala J, Huang Y, Pollock NK, Parikh S, Raed A, et al. Race/ethnicity-specific association of Vitamin D and global DNA methylation: Cross-sectional and interventional findings. PLoS ONE. 2016;11(4).
75. Zhu H, Wang X, Shi H, Su S, Harshfield GA, Gutin B, et al. A genome-wide methylation study of severe vitamin d deficiency in African American adolescents. J Pediatr. 2013;162(5):1004-9.e1.
76. Nair-Shalliker V, Dhillon V, Clements M, Armstrong BK, Fenech M. The association between personal sun exposure, serum vitamin D and global methylation in human lymphocytes in a population of healthy adults in South Australia. Mutat Res Fundam Mol Mech Mutagen. 2014;765:6-10.
77. Ashktorab H, Nguza B, Fatemi M, Nouraie M, Smoot DT, Schäffer AA, et al. Case-control study of vitamin D, dickkopf homolog 1 (DKK1) gene methylation, VDR gene polymorphism and the risk of colon adenoma in African Americans. PLoS ONE. 2011;6(10).
78. Jiang X, Yan J, West AA, Perry CA, Malysheva OV, Devapatla S, et al. Maternal choline intake alters the epigenetic state of fetal cortisol-regulating genes in humans. FASEB J. 2012;26(8):3563-74.
79. Shin W, Yan J, Abratte CM, Vermeylen F, Caudill MA. Choline intake exceeding current dietary recommendations preserves markers of cellular methylation in a genetic subgroup of folate-compromised men. J Nutr. 2010;140(5):975-80.

80. McClelland R, Christensen K, Mohammed S, McGuinness D, Cooney J, Bakshi A, et al. Accelerated ageing and renal dysfunction links lower socioeconomic status and dietary phosphate intake. Aging (Albany NY). 2016.
81. Shimazu T, Asada K, Charvat H, Kusano C, Otake Y, Kakugawa Y, et al. Association of gastric cancer risk factors with DNA methylation levels in gastric mucosa of healthy Japanese: a cross-sectional study. Carcinogenesis. 2015;36(11):1291-8.
82. Khulan B, Cooper WN, Skinner BM, Bauer J, Owens S, Prentice AM, et al. Periconceptional maternal micronutrient supplementation is associated with widespread gender related changes in the epigenome: A study of a unique resource in the Gambia. Hum Mol Genet. 2012;21(9):2086-101.
83. Pusceddu I, Herrmann M, Kirsch SH, Werner C, Hübner U, Bodis M, et al. Prospective study of telomere length and LINE-1 methylation in peripheral blood cells: the role of B vitamins supplementation. Eur J Nutr. 2015.
84. Zhou Y, Zhao LJ, Xu X, Ye A, Travers-Gustafson D, Zhou B, et al. DNA methylation levels of CYP2R1 and CYP24A1 predict vitamin D response variation. J Steroid Biochem Mol Biol. 2014;144(PART A):207-14.
85. Kok DEG, Dhonukshe-Rutten RAM, Lute C, Heil SG, Uitterlinden AG, van der Velde N, et al. The effects of long-term daily folic acid and vitamin B-12 supplementation on genome-wide DNA methylation in elderly subjects. Clinical Epigenetics. 2015;7.
86. Rerkasem K, Rattanatanyong P, Rerkasem A, Wongthanee A, Rungruengthanakit K, Mangklabruks A, et al. Higher Alu methylation levels in catch-up growth in twenty-year-old offsprings. PLoS ONE. 2015;10(3).
87. Godfrey KM, Sheppard A, Gluckman PD, Lillycrop KA, Burdge GC, McLean C, et al. Epigenetic gene promoter methylation at birth is associated with child's later adiposity. Diabetes. 2011;60(5):1528-34.
88. Lee HS, Barraza-Villarreal A, Hernandez-Vargas H, Sly PD, Biessy C, Ramakrishnan U, et al. Modulation of DNA methylation states and infant immune system by dietary supplementation with v-3 PUFA during pregnancy in an intervention study. Am J Clin Nutr. 2013;98(2):480-7.
89. Amarasekera M, Noakes P, Strickland D, Saffery R, Martino DJ, Prescott SL. Epigenome-wide analysis of neonatal CD4+ T-cell DNA methylation sites potentially affected by maternal fish oil supplementation. Epigenetics. 2014;9(12):1570-6.
90. Voisin S, Almén MS, Moschonis G, Chrousos GP, Manios Y, Schiöth HB. Dietary fat quality impacts genome-wide DNA methylation patterns in a cross-sectional study of Greek preadolescents. Eur J Hum Genet. 2015;23(5):654-62.
91. Lind MV, Martino D, Harsløf LBS, Kyjovska ZO, Kristensen M, Lauritzen L. Genome-wide identification of mononuclear cell DNA methylation sites potentially affected by fish oil supplementation in young infants: A pilot study. Prostaglandins Leukotrienes Essent Fatty Acids. 2015;101:1-7.
92. Gómez-Uriz AM, Goyenechea E, Campión J, De Arce A, Martinez MT, Puchau B, et al. Epigenetic patterns of two gene promoters (TNF-α and PON) in stroke considering obesity condition and dietary intake. J Physiol Biochem. 2014;70(2):603-14.
93. Brøns C, Jacobsen S, Nilsson E, Rönn T, Jensen CB, Storgaard H, et al. Deoxyribonucleic acid methylation and gene expression of PPARGC1A in human muscle is influenced by high-fat overfeeding in a birth-weight-dependent manner. J Clin Endocrinol Metab. 2010;95(6):3048-56.
94. Gillberg L, Jacobsen SC, Rönn T, Brøns C, Vaag A. PPARGC1A DNA methylation in subcutaneous adipose tissue in low birth weight subjects - Impact of 5 days of high-fat overfeeding. Metab Clin Exp. 2014;63(2):263-71.

95. Gillberg L, Perfilyev A, Brøns C, Thomasen M, Grunnet LG, Volkov P, et al. Adipose tissue transcriptomics and epigenomics in low birthweight men and controls: role of high-fat overfeeding. Diabetologia. 2016;59(4):799-812.
96. Jacobsen SC, Gillberg L, Bork-Jensen J, Ribel-Madsen R, Lara E, Calvanese V, et al. Young men with low birthweight exhibit decreased plasticity of genome-wide muscle DNA methylation by high-fat overfeeding. Diabetologia. 2014;57(6):1154-8.
97. Hermsdorff HH, Mansego ML, Campión J, Milagro FI, Zulet MA, Martínez JA. TNF-alpha promoter methylation in peripheral white blood cells: Relationship with circulating TNFα, truncal fat and n-6 PUFA intake in young women. Cytokine. 2013;64(1):265-71.
98. Milagro FI, Gómez-Abellán P, Campión J, Martínez JA, Ordovás JM, Garaulet M. CLOCK, PER2 and BMAL1 DNA methylation: Association with obesity and metabolic syndrome characteristics and monounsaturated fat intake. Chronobiol Int. 2012;29(9):1180-94.
99. Ma Y, Smith CE, Lai CQ, Irvin MR, Parnell LD, Lee YC, et al. The effects of omega-3 polyunsaturated fatty acids and genetic variants on methylation levels of the interleukin-6 gene promoter. Mol Nutr Food Res. 2016;60(2):410-9.
100. Ma Y, Follis JL, Smith CE, Tanaka T, Manichaikul AW, Chu AY, et al. Interaction of methylation-related genetic variants with circulating fatty acids on plasma lipids: A meta-analysis of 7 studies and methylation analysis of 3 studies in the Cohorts for Heart and Aging Research in Genomic Epidemiology consortium. Am J Clin Nutr. 2016;103(2):567-78.
101. Samblas M, Milagro FI, Gomez-Abellan P, Martinez JA, Garaulet M. Methylation on the Circadian Gene BMAL1 Is Associated with the Effects of a Weight Loss Intervention on Serum Lipid Levels. J Biol Rhythms. 2016.
102. Peters MJ, Joehanes R, Pilling LC, Schurmann C, Conneely KN, Powell J, et al. The transcriptional landscape of age in human peripheral blood. Nat Commun. 2015;6:8570.
103. Lee HS, Barraza-Villarreal A, Biessy C, Duarte-Salles T, Sly PD, Ramakrishnan U, et al. Dietary supplementation with polyunsaturated fatty acid during pregnancy modulates DNA methylation at IGF2/H19 imprinted genes and growth of infants. Physiol Genomics. 2014;46(23):851-7.
104. Piyathilake CJ, Robinson CB, Cornwell P. A practical approach to red blood cell folate analysis. Anal Chem Insights. 2007;2:107-10.
105. De Bruyn E, Gulbis B, Cotton F. Serum and red blood cell folate testing for folate deficiency: new features? Eur J Haematol. 2014;92(4):354-9.



PART B

Homocysteine and DNA methylation

CHAPTER 4

Homocysteine levels associate with subtle changes in leukocyte DNA methylation: an epigenome-wide analysis

Pooja R. Mandaviya*, Dylan Aïssi*, Koen F. Dekkers*, Roby Joehanes*, Silva Kasela*, Vinh Truong*, Lisette Stolk, Diana van Heemst, M. Arfan Ikram, Jan Lindemans, P. Eline Slagboom, David-Alexandre Trégouët, André G. Uitterlinden, Chen Wei, Phil Wells, BIOS consortium, France Gagnon**, Marleen MJ van Greevenbroek**, Bastiaan T. Heijmans**, Lili Milani**, Pierre-Emmanuel Morange**, Joyce B.J. van Meurs**, Sandra G. Heil**

**These authors contributed equally as first authors*
***These authors contributed equally as senior authors*

Epigenomics. 2017 Nov;9(11):1403-1422

ABSTRACT

Aim: Homocysteine is a sensitive marker of one-carbon metabolism. Higher homocysteine levels have been associated with global DNA hypomethylation. We investigated the association between plasma homocysteine and epigenome-wide DNA methylation in leukocytes.

Methods: Methylation was measured using Illumina 450k arrays in 2,035 individuals from 6 cohorts. Homocysteine-associated differentially methylated positions (DMPs) and regions (DMRs) were identified using meta-analysis.

Results: Three DMPs cg21607669 (*SLC27A1*), cg26382848 (*AJUBA*) and cg10701000 (*KCNMA1*) at chromosome 19, 14 and 10, respectively, were significantly associated with homocysteine. In addition, we identified 68 homocysteine-associated DMRs, the most significant of which was a 1.8 kb spanning domain (*TNXB/ATF6B*) at chromosome 6.

Conclusion: We identified novel epigenetic loci associated with homocysteine levels, of which the specific role needs to be further validated.

SUMMARY POINTS

1) Plasma homocysteine is significantly associated with a modest number of differentially methylated positions (DMPs) in leukocyte DNA.
2) 68 DMRs related to 114 genes were significantly associated with plasma homocysteine. These genes were involved in folate biosynthesis, glycosaminoglycan biosynthesis, arachidonic acid metabolism and glycerophospholipid metabolism.

INTRODUCTION

Homocysteine (Hcy), an intermediate of the one-carbon metabolism, is an independent risk factor of diseases like osteoporotic fractures [1] and cardiovascular disorders [2]. In addition, Hcy is often used as sensitive marker of folate- and vitamin B12 deficiency. Recent intervention trials were not able to show that lowering of plasma Hcy by B-vitamin intervention led to a reduction in disease risk such as myocardial infarction [3-5] and osteoporotic fractures [6]. Identification of the underlying pathophysiological mechanism of Hcy is warranted to understand its role as biomarker of several disease states.

Hcy is formed by a demethylation reaction from methionine. In the methionine cycle, s-adenosylmethionine (SAM) donates its methyl group to the DNA resulting in s-adenosylhomocysteine (SAH) as a by-product and a potent inhibitor of this methyltransferase reaction. SAH is hydrolyzed to Hcy in a reversible manner. Elevated Hcy in a condition known as hyperhomocysteinemia (HHcy) is associated to and increased SAH level, which is linked to alterations in DNA methylation [7-9]. This is one of the possible mechanisms in Hcy-related pathologies and the rational to study the effect of Hcy on genome-wide DNA methylation.

Hcy is a sensitive marker of one-carbon metabolism and will increase when remethylation or transsulfuration is impaired. Shortage of folate, vitamin B12 or vitamin B6 result in elevated Hcy levels. Although folate levels are important, we believe it is not necessary to correct for folate in our analysis as Hcy is strongly correlated to folate and belongs to the same pathway. This would have cancelled out the effect of Hcy itself on DNA methylation. The general hypothesis of this study was that Hcy is associated with changes in DNA methylation irrespective of the cause.

Normal circulating Hcy levels ranges from 8 to 10 µmol/L for women and 10 to 12 µmol/L for men [10]. Mild or moderate HHcy is referred to Hcy concentrations between 15-30 µmol/L, whereas intermediate and severe HHcy is referred to Hcy concentrations between 30-100 µmol/L and >100 µmol/L, respectively [10-12]. HHcy has been shown to be associated with both global- and gene-specific DNA hypomethylation, which was recently reviewed by us [13].. In candidate gene studies of animal models, genes whose methylation have shown association with Hcy are the hypomethylation of imprinted genes (e.g. *IGF2/H19* locus) [14, 15], cardiac *Trp53* [16], leukocytic *mTERT* promoter [17], brain *Nr3c1* exon region [18], brain *Ppp2r1a* [19] and liver 5' portion of *Soat2* [20], and hypermethylation of the aortic *Nr3c1* promoter [21] and liver *Fads2* promoter [22]. In human studies, HHcy has been shown to be associated with altered methylation of the *RFC1* gene [23]. For patients with vascular disorders, HHcy has shown association with hypomethylation of the leukocytic *hTERT* promoter [17], leukocytic *SHC1* promoter [24], punch aortic *NOS2* proximal promoter [25], and hypermethylation of the blood ESR1 promoter [26] and internal mammary artery *NOS2* proximal promoter [25]. For patients with colorectal, esophageal and

breast cancer, HHcy has shown association with hypermethylation of *ESR1* promoter [27] and exon 1 [28], *MLH1* promoter [29], *RARB* promoter exon [28], *BRCA1* and *RASSF1* promoter [30].

In animal models, HHcy was also shown to associate with global hypomethylation [18, 31-33]. Human studies included mostly small-scale studies in renal or vascular patients with HHcy in which global DNA hypomethylation was observed [10, 34, 35]. Most studies showed significant associations of Hcy with DNA methylation, however, also null associations have been found [36, 37]. Hence, the association between Hcy and DNA methylation is complex, contradictory and remains to be fully understood. Mild HHcy can occur due to either genetic or environmental factors such as low intake of B vitamins. Regardless of the cause, Hcy itself is thought to be the most prominent key player in altering DNA methylation patterns and associated disease risks.

In the current manuscript, we explored the association of Hcy in relation to genome-wide CpG site level in leukocytes to assess its role in DNA methylation. In this study, we performed a meta-analysis in 2,035 individuals of 6 individual cohorts using the high-resolution Infinium Illumina 450k BeadChip arrays.

MATERIALS AND METHODS

Cohorts

Six cohorts participated in the analysis, Rotterdam study [38] (RS, n=700), Leiden Longevity Study [39] (LLS, n=550), Cohort on Diabetes and Atherosclerosis Maastricht [40] (CODAM, n=186), MARseille THrombosis Association study [41] (MARTHA, n=293), Estonian Biobank [42] (EGCUT, n=96), and French-Canadian family study on Factor V Leiden (F5L) thrombophilia [41] (F5L, n=210). This added up to a total sample size of 2,035 for the meta-analysis. The ethics review committees approved each study at their relevant organizations. Characteristics of each study are provided in Table 1 and Text S1.

Table 1: Characteristics of cohorts included in the Hcy meta-analysis.

Cohort (Country)	N	Type (Pathophysiology)	Women (%)	Hcy‡ [Median (Range)]	HHcy (%)	Method	Total Precision [C.V.%]	Time between blood collecting & freezing	Age [Mean±SD]	Gran [Mean±SD]	Lym [Mean±SD]	Mono [Mean±SD]
RS (NL)	700	Prospective, Population-based[A]	54.7	11.5 (5.9-35.4)	19.4	LC-MS/MS	5.5	ASAP	59.8±8.0	4.2±1.5	2.4±0.7	0.4±0.4
LLS (NL)	550	Prospective, Family-based[A]	51.8	12.2 (5.3-31.6)*	18.2	Competitive Immunoassay	NA	ASAP	58.7±6.6	4.4±1.3	2.0±0.6	0.4±0.1
CODAM (NL)	186	Prospective, Observational[B]	54.8	10.5 (5.6-29.9)	11.3	LC-MS/MS	<4.0	<2 hrs	65.3±7.0	0.3±0.1	0.1±0.1[W] 0.3±0.1[X] 0.2±0.1[Y] 0.1±0.0[Z]	0.1±0.1
MARTHA (FR)	293	Retrospective[C]	80.3	10.3 (3.4-33.8)	10.2	Competitive Immunoassay	5.0	<2 hrs	43.5±14.3	4.4±1.6	1.9±0.6	0.3±0.1
EGCUT (EE)	96	Population-based[A]	50.0	13.3 (6.3-34.0)*	32.3	LC with Photometric detection	NA	No freezing	52.4±23.7	3.8±1.3	2.0±0.7	0.5±0.2
F5L (CA)	210	Pedigree-based[A,D]	52.4	8.3 (3.7-28.2)	3.3	Fluorescence Polarization Immunoassay	5.2	1-3 hrs	39.8±16.8	0.6±0.1	0.0±0.0[W] 0.2±0.1[X] 0.1±0.1[Y] 0.1±0.0[Z]	0.1±0.0

*Non-fasting; ‡Hcy concentrations in µmol/L; NA=Not Available, ASAP (As soon as possible), HHcy=Hcy concentrations >15 µmol/L
Gran, Lym, Mono (Measured cell counts in 10^9/L), [W] CD8T, [X] CD4T, [Y] NK, [Z] Bcell (Houseman estimated percentage cell counts)
Pathophysiology: [A]Mainly healthy, [B]Mildly increased DM2/CVD risk factors, [C]Patients with Venous Thromboembolism, [D]Probands with VTE

Homocysteine measurements

Total Hcy was measured in EDTA plasma. Measurements in non-fasting state were performed in the LLS and EGCUT study, while rest of the studies measured Hcy in fasting state. Details of each method is depicted in Table 1. Since the absolute values of Hcy in all studies were skewed to the right, logarithmic transformation of base 10 was used to achieve normal distribution. For a sample to be considered as an outlier, an outlier detection strategy [43] was used with four parameters of the Hcy log10 transformed data: quartile 1 (Q1), quartile 3 (Q3), interquartile range (IQR) and a multiplying factor (M). The lower end outliers (LO) were detected by the formula, LO=Q1–(M*IQR), and upper end outliers (UO) were detected using UO=Q3+(M*IQR). Any value less than or equal to LO and greater than or equal to UO was considered as an extreme outlier, using the multiplying factor 3. A logarithmic transformation of base 10 and removal of outliers showed a clear improvement towards normal distribution [Figure S1].

DNA methylation assessment

Genomic DNA was extracted from whole blood and was bisulfite converted. Infinium Illumina HumanMethylation 450k BeadChip arrays (Illumina Inc., San Diego, USA) were used to perform DNA methylation profiling according to the manufacturers' protocol. This assay covers methylation status of a wide range of CpGs at more than 450,000 sites at regions of CpG islands, shores and shelves. The sites annotate to various gene regions like promoters, enhancers, gene bodies, exons and UTRs [44]. Standard beta values were considered for the methylation analysis in all studies. These contain values ranging from 0 to 1, representing the percentage of methylation. Beta values are calculated from the methylated (M) and unmethylated (U) probe intensities using the formula of $\beta=M/(M+U+\alpha)$, with the default alpha (α) of 100. These values were normalized using methods, such as DASEN [45] or SWAN [46]. Details of DNA methylation assessment, pre-processing methods, and sample and probe quality control used by each cohort are provided in Text S1 and Table S1.

Differentially Methylated Positions (DMPs)

Statistical analysis to identify DMPs was performed using R programming scripts, which were send around to each cohort and run by each cohort separately with log10 transformed Hcy as the exposure variable and DNA methylation as an outcome. This association analysis was performed using linear mixed models of the lme4 package in R. The association model was adjusted for technical covariates and biological covariates like age, sex and differential white blood cell (WBC) counts. Array number and position on array were used as technical covariates, and were treated as random effects. For cohorts that did not contain measured WBC counts, the counts were

inferred using the Houseman method as implemented in the minfi package [47] [Table 1].

Association analyses provided summary statistics from each study. A fixed effect meta-analysis was performed in 2,035 individuals of 6 cohorts using the classical approach of the METAL software [48] that combines effect size estimates and standard errors of each study summary statistics. Probes with SNPs at single base extension site were removed to avoid confounding and spurious signals. Probes with improper binding were also removed to avoid co-hybridization with alternate homologous sequences [49]. In addition, we removed probes that were present in less than 4 studies since each cohort had different probe exclusions. This resulted in 468,108 probes remaining. Benjamini-Hochberg (BH) method was used to define statistical significance for a false discovery rate (FDR) less than 0.05. The genomic inflation factor (λ) was calculated [50] and cohort heterogeneity was considered by taking into account the heterogeneity I^2 parameter [48]. CpGs for nearby genes were annotated using the Genomic Regions Enrichment of Annotations Tool (GREAT) [51]. This assigns extension up to 5 kb upstream and 1 kb downstream to a basal regulatory region from its transcription start site and a maximum up to 1 Mb extension distance, as defined by UCSC [52].

IGF2/H19 locus

We also looked up for the three DMRs of IGF2/H19 locus at chromosome 11, which were previously related with Hcy [14, 15]. The seven 450k array CpGs underlying the three DMRs of this locus were identified using literature [36, 53, 54]. Multiple testing using the Bonferroni method was applied on these 7 CpGs to test for their significance with Hcy.

Enrichment of previously found folate-associated CpGs

Because Hcy is associated with folate, we additionally focused our analysis on the previously found 443 CpGs in cord blood of newborns associated with maternal plasma folate during pregnancy [55]. In order to check for their significant enrichment, we compared their p-values with the p-values of randomly selected 443 CpGs from the array, using the Fisher’s exact test.

Differentially Methylated Regions (DMRs)

DMRs were identified using the python library of the comb-p software [56]. Absolute p-values from the meta-analysis of DMP analysis were given as input to combine them [57] adjacently at varying distance lags of 50 base pairs (bp) in a sliding-window of 500 bp, in order to calculate auto-correlation between them. As weighted according to this auto-correlation, each p-value is adjusted according to its adjacent p-values. An FDR correction using the BH method is then calculated on these adjusted p-values. A peak-finding algorithm then finds enrichment regions on these FDR corrected adjusted p-

values, and using Stouffer-Liptak correction as implemented in comb-p [56, 58, 59], it then assigns new p-values to the identified regions using the original absolute p-values from the meta-analysis. Finally, these assigned new p-values of the of the identified regions are multiple tested using the Sidak correction [60], where the defined statistical significance for the identified regions was Sidak corrected P less than 0.05. The number of tests for a region in the Sidak correction is the total bases covered by input probes divided by the region size [56].

Differentially Methylated Regions (DMRs): Pathway Analysis

For the pathway analysis, we annotated the regions for genes using the tool GREAT [51]. We used WEB-based GEne SeT AnaLysis Toolkit (WebGestalt) [61] online resource for pathway analysis. KEGG database were used for enrichment. BH method in WebGestalt was used to define statistical significance for an FDR less than 0.05.

RESULTS

Cohort characteristics

Meta-analysis comprised of 6 cohorts with a total of 2,035 individuals. All individuals were of European ancestry. The study design was mainly population based, but also included cohorts with increased risk of diabetes, cardiovascular disorders and venous thromboembolism. Study characteristics for all the participating cohorts are given in Table 1.

Differentially Methylated Positions (DMPs)

Meta-analysis on 468,108 CpGs was done on the summary statistic of the 6 cohorts [Figure S2] comprising of 2,035 individuals. The p-values of this analysis showed a genomic inflation factor of 1.2 [Figure S3]. Three DMPs [Table 2A, Figure 1] were significantly associated with Hcy (FDR less than 0.05). All 3 DMPs showed 1% increase in methylation with each unit increase in the 10 logarithmic Hcy value. The DMP cg21607669 located near the gene *SLC27A1* at chromosome 19 had the lowest p-value of 7.71E-08 (FDR=0.04). The 2 other DMPs were cg26382848 (nearest gene *AJUBA*) located at chromosome 14 with p-value 1.67E-07 (FDR=0.04) and cg10701000 (nearest gene *KCNMA1*) located at chromosome 10 with p-value 2.38E-07 (FDR=0.04). These 3 DMPs showed no substantial cohort heterogeneity I^2 (I^2 less than 35). The direction of effects were mostly similar between all studies [Figure 2A, 2B & 2C].

IGF2/H19 locus

We did a lookup for CpGs at the 3 DMRs of the IGF2/H19 gene (DMR0, DMR2, H19-DMR3), that have been previously studied in humans [36, 53, 54]. Data from our 450k arrays contained 7 CpGs at these DMRs: 2 at DMR0, 4 at DMR2 and 1 at H19-DMR3.

However, none of the 7 CpGs showed an association with Hcy with a Bonferroni cut off of 7.14E-03 [Table 2B, Figure S5].

Enrichment of previously found folate-associated CpGs

For the previously found 443 CpGs in cord blood of newborns associated with maternal plasma folate during pregnancy [55], we compared their nominal p-values with 443 randomly selected CpGs in the array using 100 permutations. Based on our significant threshold (P less than 0.05), 20 of the 443 folate-related CpGs met the threshold as compared to the 32 of the 443 randomly selected CpGs in the array. Using the fisher's exact test, we found no significant enrichment (p-value=0.97) in the folate-related CpGs.

Differentially Methylated Regions (DMRs)

We identified 68 DMRs significant at Sidak P less than 0.05 [Table 3, Figure 3, Figure S4]. The most significant was the DMR at chromosome 6 (spanning a region of 1,8 Kb) with a p-value of 4.34E-24 (Sidak P = 1.12E-21) containing 55 CpGs. CpGs within this region were annotated to gene *TNXB* and *ATF6B*. Pathway analysis on the 114 genes annotated to the 68 DMRs showed 14 significant pathways [Table 4]. Top 5 pathways were metabolic pathways, folate biosynthesis, glycosaminoglycan biosynthesis – heparan sulfate, phagosome and MAPK signaling pathway. Furthermore, results of the GO enrichment analysis showed many biological processes related to embryogenesis and development [Table 5].

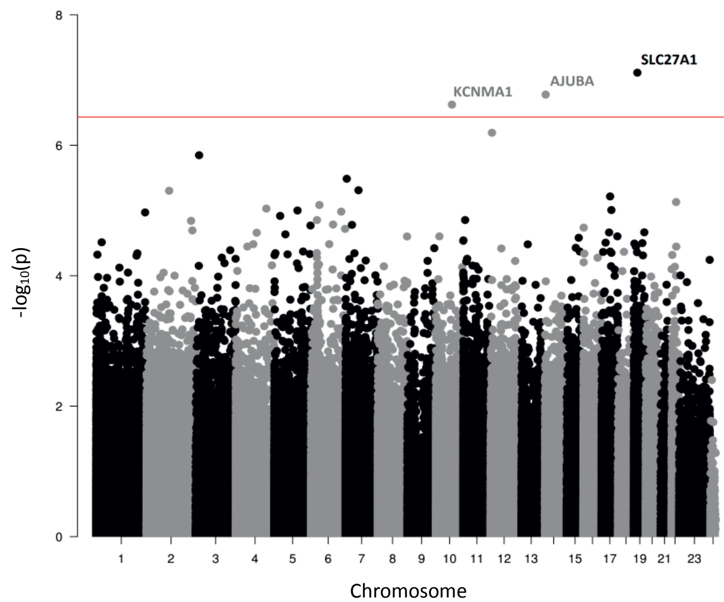


Figure 1. Manhattan plot showing the association between homocysteine and genome-wide DNA methylation in 2,035 samples, with 3 significant DMPs at chromosomes 10, 14 and 19, at FDR less than 0.05 (red line). Nearest genes for these 3 DMPs are reported.

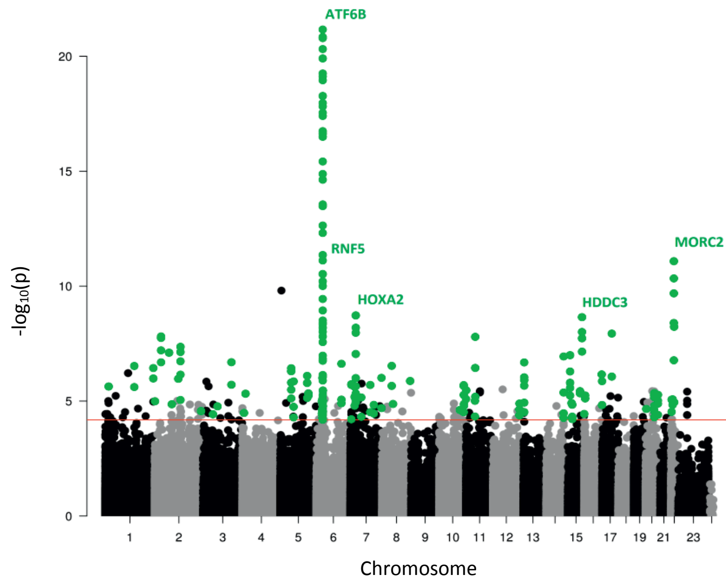
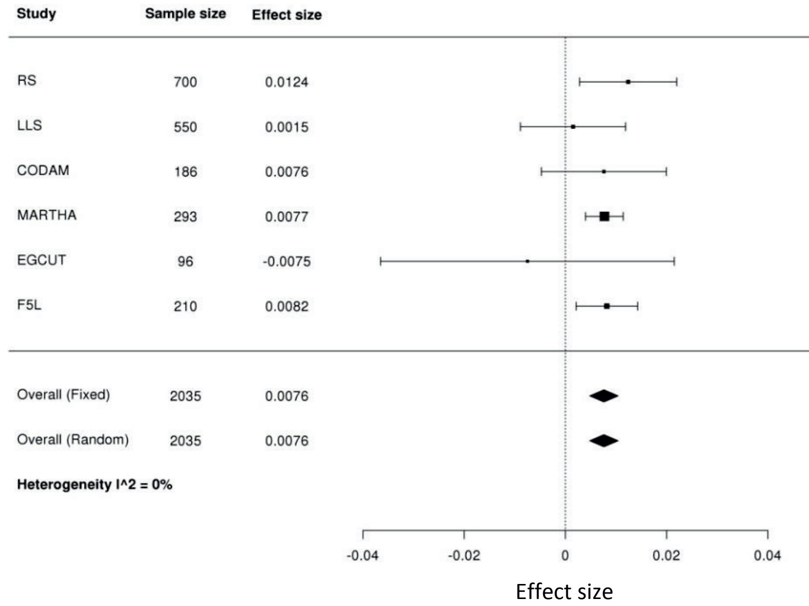


Figure 3. Manhattan plot showing the association between homocysteine and genome-wide DNA methylation in 2,035 samples, with 68 significant DMRs, at FDR less than 0.05 (red line) of autocorrelation adjusted p-values in comb-p. Nearest genes for the top 5 DMRs are reported.



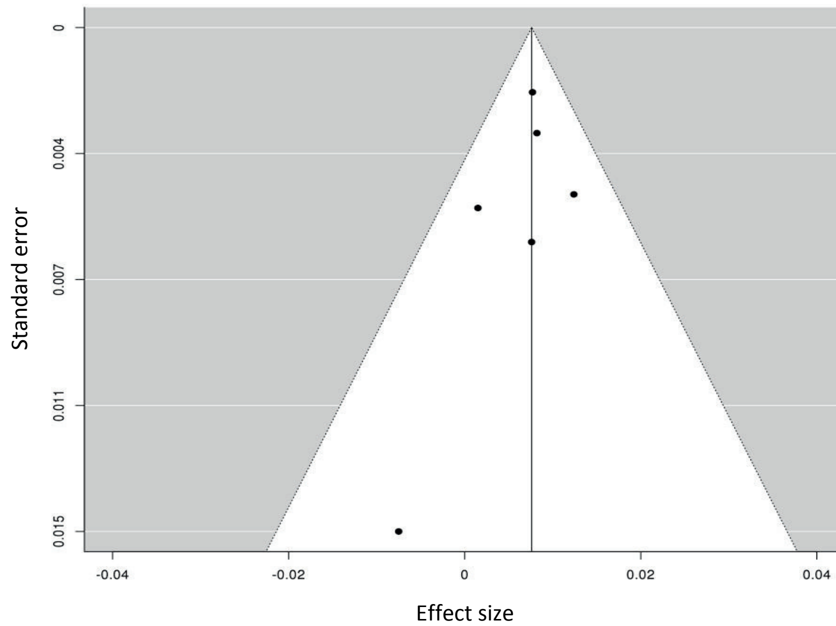
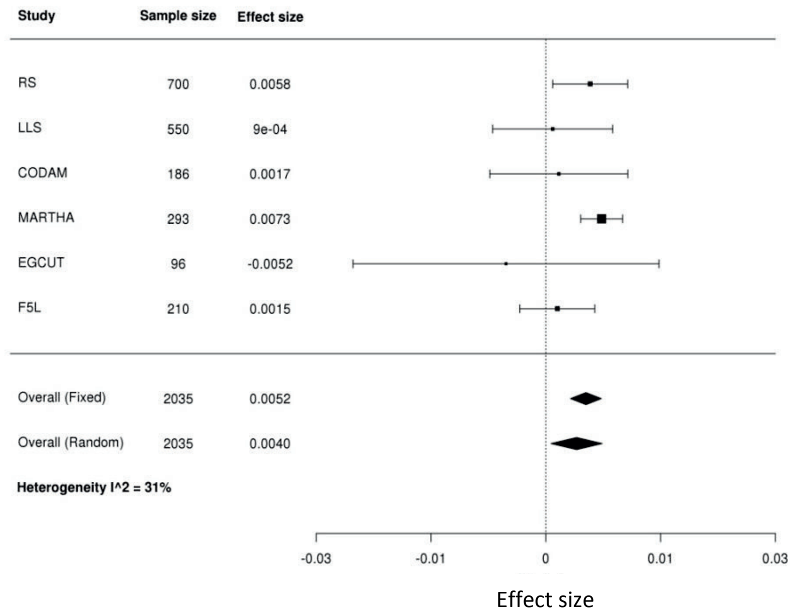


Figure 2A. Forest plot showing the association between homocysteine and the first significant DMP cg21607669 across the 6 cohorts.



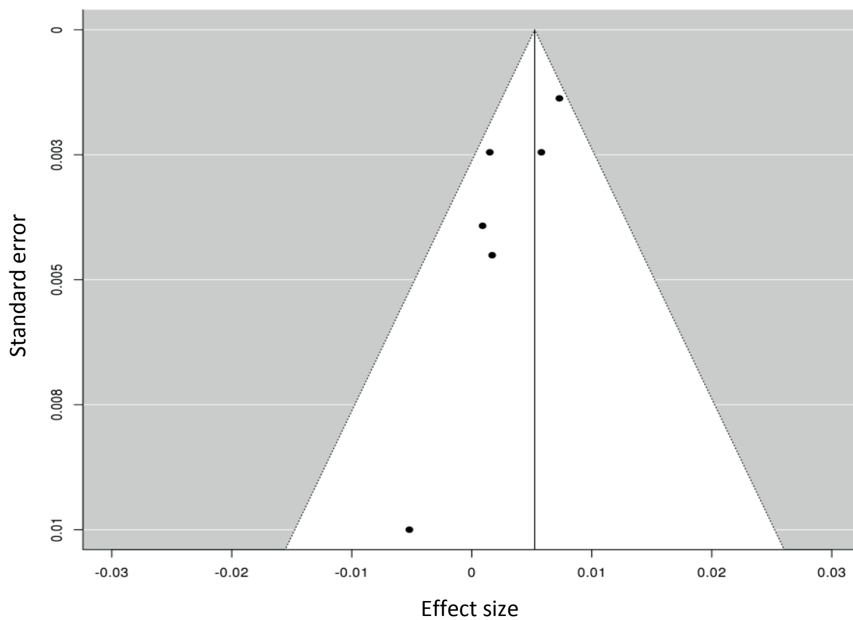


Figure 2B. Forest plot showing the association between homocysteine and the second significant DMP cg26382848 across the 6 cohorts.

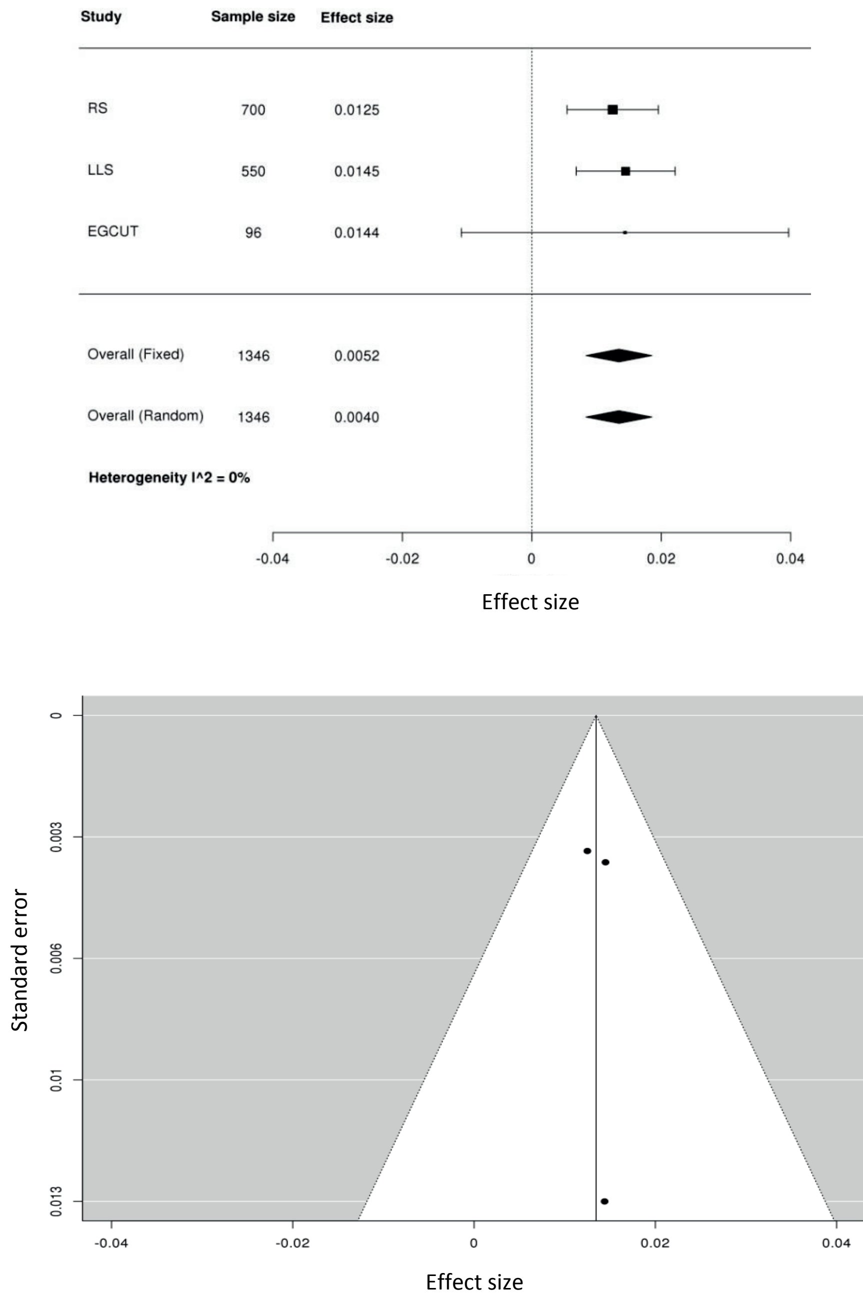


Figure 2C. Forest plot showing the association between homocysteine and the third significant DMP cg10701000 across the 3 cohorts.

Table 2A. Homocysteine-associated DMPs

Rank	CpG	N	Effect	StdErr	Pvalue	FDR	HetISq	Nearby Gene (±bp)	Chr	Bp	Location
1	cg21607669	2,035	0.01	0.001	7.71E-08	0.036	0	SLC27A1 (+40)	19	17581292	Promoter
2	cg26382848	2,035	0.01	0.001	1.67E-07	0.037	31.1	AJUBA (-39)	14	23451889	Promoter
3	cg10701000	1,346	0.01	0.003	*2.38E-07*	0.037	0	KCNMA1 (-143,626), DLG5 (+145,258)	10	79541025	-

Effect: Beta coefficients based on log transformed Hcy, HetISq: Heterogeneity I^2 parameter
Bp: Base pair location based on Illumina annotation, Location: Based on Illumina annotation, derived from UCSC
FDR threshold = 0.05

Table 2B. Homocysteine-associated DMPs at the 3 IGF2/H19 DMR regions at chromosome 11

DMR	CpG	N	Effect	StdErr	Pvalue	Bonferroni	HetISq	Bp
H19-DMR3	cg22259242	1532	0.0032	0.0048	0.5014	No	8.6	2021243
DMR2	cg13165070	1849	0.0041	0.011	0.713	No	0	2154113
DMR2	cg11717189	1753	-0.0079	0.0143	0.5824	No	0	2154132
DMR2	cg07096953	2035	-0.002	0.0106	0.847	No	0	2154255
DMR2	cg02613624	2035	0.0084	0.0072	0.2421	No	32.9	2154386
DMR0	cg00273464	2035	0.024	0.0094	0.0107	No	54.2	2170412
DMR0	cg17665927	2035	0.0054	0.0098	0.584	No	30.3	2170443

Effect: Beta coefficients based on log transformed Hcy, HetISq: Heterogeneity I^2 parameter
Bp: Base pair location based on Illumina annotation, Bonferroni threshold = 7.14E-03

Table 3. Homocysteine-associated DMRs

Rank	Chr	Start	End	Bp length	# CpGs	P	Sidak-P	Mean effect size	Annotated genes	Location
1	6	32063394	32065212	1818	55	4.34E-24	1.12E-21	-0.04	TNXB (-50398), ATF6B (+31714)	-
2	22	31317764	31318547	783	12	1.28E-11	7.67E-09	-0.04	MORC2 (+46031), OSBP2 (+227363)	Enhancer
3	6	32145146	32146780	1634	29	6.53E-10	1.87E-07	-0.01	AGPAT1 (-1106), RNF5 (-168)	Promoter
4	7	27142100	27142811	711	15	2.70E-09	1.78E-06	-0.02	HOXA2 (-26)	-
5	15	91473059	91473570	511	9	2.59E-09	2.37E-06	-0.02	HDDC3 (+2461), MAN2A2 (+25895)	Promoter
6	22	32598479	32598717	238	4	5.87E-09	1.16E-05	-0.03	RFPL2 (+866)	-
7	17	48473757	48474100	343	4	1.16E-08	1.59E-05	-0.02	LRRC59 (+985)	-
8	2	31806352	31806899	547	6	1.94E-08	1.66E-05	-0.03	XDH (-169045), MEMO1 (+429000)	-
9	11	45827260	45827696	436	6	1.61E-08	1.73E-05	-0.01	CRY2 (-41479), SLC35C1 (+1521)	-
10	2	129659316	129659947	631	6	5.12E-08	3.80E-05	-0.04	HS6ST1 (-583481)	-
11	15	31515750	31516482	732	9	1.34E-07	8.55E-05	-0.03	KLF13 (-102942), TRPM1 (-62640)	Enhancer
12	10	123355268	123356042	774	5	1.62E-07	9.81E-05	-0.06	FGFR2 (+2262), WDR11 (+744968)	-
13	2	72079276	72079610	334	6	7.86E-08	1.10E-04	-0.04	CYP26B1 (+295724), DYSF (+385611)	-
14	20	3643863	3644193	330	5	8.80E-08	1.25E-04	-0.02	GFRA4 (+18)	-
15	14	106321551	106322430	879	7	2.89E-07	1.54E-04	0.01	TMEM121 (+329051)	-
16	6	126080132	126080724	592	4	2.57E-07	2.03E-04	-0.02	NCOA7 (-21879), HEY2 (+9702)	-
17	3	142666108	142666477	369	4	2.04E-07	2.59E-04	-0.04	PCOLCE2 (-58248), PAQR9 (+15885)	-
18	1	242220475	242220926	451	4	3.66E-07	3.80E-04	-0.04	MAP1LC3C (-58326), PLD5 (+467119)	-
19	5	54281198	54281734	536	8	5.68E-07	4.96E-04	-0.02	ESM1 (+25)	Enhancer
20	16	87978839	87979203	364	3	6.93E-07	8.91E-04	-0.04	BANP (-24603), CA5A (-8899)	-
21	5	134735544	134735915	371	8	8.01E-07	1.01E-03	-0.02	H2AFY (-418)	-
22	2	118616155	118616577	422	5	1.08E-06	1.20E-03	-0.04	INSIG2 (-229684), DDX18 (+44140)	-
23	16	89689811	89690263	452	3	1.37E-06	1.42E-03	-0.04	SPATA33 (-34173), DPEP1 (+3037)	-
24	13	23412250	23412623	373	4	1.22E-06	1.53E-03	-0.06	SGCG (-342654)	-
25	6	32805398	32805693	295	6	1.21E-06	1.91E-03	-0.02	TAP2 (+958)	Promoter

Rank	Chr	Start	End	Bp length	# CpGs	P	Sidak-P	Mean effect size	Annotated genes	Location
26	7	4764845	4765313	468	4	2.01E-06	2.01E-03	-0.03	AP5Z1 (-50174), FOXK1 (+43139)	Promoter, Enhancer
27	13	23310188	23310676	488	6	2.36E-06	2.26E-03	-0.04	SGCG (-444659)	-
28	8	48744176	48744603	427	5	2.19E-06	2.40E-03	0.01	CEBPD (-92742), PRKDC (+128353)	-
29	7	21209338	21209782	444	4	2.38E-06	2.51E-03	-0.03	SP8 (-383055), SP4 (-258092)	Enhancer
30	8	141359539	141359787	248	4	1.34E-06	2.53E-03	-0.04	TRAPPC9 (+109015), C8orf17 (+416247)	Enhancer
31	7	99724112	99724444	332	4	1.99E-06	2.80E-03	-0.01	MBLAC1 (-39)	Promoter
32	11	396686	397078	392	3	3.42E-06	4.07E-03	-0.06	PKP3 (+2665), SIGIRR (+18094)	-
33	20	62367698	62368257	559	8	5.46E-06	4.56E-03	0.02	SLC2A4RG (-3236), LIME1 (-17)	Promoter
34	4	15704393	15704845	452	8	4.79E-06	4.95E-03	-0.01	BST1 (+46)	-
35	15	81426347	81426670	323	9	3.71E-06	5.36E-03	0.03	MESDC2 (-144290), IL16 (-91232)	-
36	5	140762229	140762583	354	4	4.62E-06	6.09E-03	-0.03	PCDHGA7 (-61)	-
37	1	17634543	17634717	174	4	2.33E-06	6.25E-03	-0.02	PADI4 (-62)	-
38	16	4802600	4802991	391	4	5.69E-06	6.79E-03	0.01	ZNF500 (+14423), NUDT16L1 (+59101)	-
39	20	43883307	43883747	440	4	6.81E-06	7.22E-03	-0.02	SLPI (-322)	-
40	22	25160033	25160407	374	6	8.42E-06	1.05E-02	-0.03	PIWIL3 (+10463), GGT1 (+161052)	-
41	7	54732478	54732752	274	2	6.94E-06	1.18E-02	-0.03	SEC61G (+95052), VSTM2A (+122597)	-
42	7	27187269	27187692	423	9	1.23E-05	1.35E-02	-0.01	HOXA5 (-4194), HOXA6 (-88)	-
43	15	42371511	42371968	457	7	1.49E-05	1.51E-02	-0.02	PLA2G4E (-28352), PLA2G4D (+15012)	Enhancer
44	2	1480789	1481098	309	3	1.02E-05	1.54E-02	-0.04	TPO (+63711), PXDN (+267334)	-
45	2	86038423	86038803	380	4	1.38E-05	1.69E-02	-0.02	ATOH8 (+57596), ST3GAL5 (+77524)	Enhancer
46	15	38988533	38988861	328	4	1.23E-05	1.74E-02	-0.02	THBS1 (-884597), RASGRP1 (-131690)	-
47	8	55294536	55294883	347	6	1.34E-05	1.79E-02	-0.01	SOX17 (-75785), MRPL15 (+246940)	Enhancer
48	22	20009063	20009276	213	2	8.62E-06	1.88E-02	-0.01	ARVCF (-4839), TANGO2 (+539)	Promoter
49	11	44327869	44328155	286	2	1.26E-05	2.04E-02	-0.02	ALX4 (+3704), EXT2 (+210265)	-
50	3	72704324	72704702	378	4	1.71E-05	2.09E-02	-0.03	RYBP (-208444), SHQ1 (+193118)	Enhancer
51	10	124578209	124578545	336	4	1.60E-05	2.21E-02	-0.04	CUZD1 (+31932), DMBT1 (+258196)	-
52	5	66462293	66462663	370	3	1.78E-05	2.23E-02	-0.03	CD180 (+30149), MAST4 (+570289)	Promoter, Enhancer

Rank	Chr	Start	End	Bp length	# CpGs	P	Sidak-P	Mean effect size	Annotated genes	Location
53	20	43726431	43726766	335	3	1.64E-05	2.27E-02	-0.03	KCNS1 (+3154), STK4 (+131484)	Promoter
54	10	135092104	135092242	138	2	8.19E-06	2.74E-02	-0.02	ADAM8 (-1801)	-
55	2	233251770	233252171	401	4	2.72E-05	3.13E-02	0.02	ALPPL2 (-19582), ALPP (+8727)	Enhancer
56	7	156400711	156400991	280	4	2.02E-05	3.32E-02	-0.04	SHH (-795884), C7orf13 (+32497)	-
57	10	104196206	104196542	336	5	2.44E-05	3.34E-02	-0.02	CUEDC2 (-3956)	-
58	4	8126221	8126659	438	4	3.31E-05	3.47E-02	0	AFAP1 (-184787), ABLIM2 (+33996)	-
59	20	42955472	42955782	310	3	2.47E-05	3.66E-02	-0.02	FITM2 (-15818), R3HDML (-9999)	-
60	22	32599511	32599649	138	5	1.13E-05	3.77E-02	-0.05	RFPL2 (-116)	-
61	6	27185676	27186077	401	4	3.54E-05	4.04E-02	-0.05	PRSS16 (-29625), HIST1H2AH (+71016)	-
62	12	133000383	133000831	448	4	4.13E-05	4.23E-02	0.02	GALNT9 (-310034), MUC8 (+50119)	-
63	1	146551565	146551745	180	2	1.68E-05	4.27E-02	-0.02	PRKAB2 (+92468), NBPF12 (+177599)	-
64	6	28601269	28601520	251	14	2.43E-05	4.44E-02	-0.03	SCAND3 (-46283), TRIM27 (+290371)	-
65	7	117854304	117854635	331	3	3.48E-05	4.81E-02	-0.02	ANKRD7 (-10260), NAA38 (+30384)	Promoter
66	6	30656499	30656693	194	7	2.06E-05	4.86E-02	0.01	PPP1R18 (-924)	-
67	3	49723947	49724292	345	3	3.75E-05	4.96E-02	-0.02	RNF123 (-2812), MST1 (+2366)	-
68	16	129230	129563	333	3	3.63E-05	4.98E-02	0.01	MPG (+1141), NPRL3 (+59272)	Promoter

Table 4. KEGG pathway analysis using WebGestalt tool

Pathways	Enrichment Ratio	rawP	adjP	No. of Genes	Genes
METABOLISM					
Metabolic pathways	4.13	3.60E-05	0.0006	12	MAN2A2, GGT1, ALPPL2, GALNT9, ALPP, AGPAT1, TPO, BST1, ST3GAL5, EXT2, XDH, PLA2G4E
Folate biosynthesis	70.64	0.0004	0.0034	2	ALPPL2, ALPP
Glycosaminoglycan biosynthesis - heparan sulfate	29.89	0.002	0.0113	2	HS6ST1, EXT2
Arachidonic acid metabolism	13.17	0.0102	0.0217	2	GGT1, PLA2G4E
Glycerophospholipid metabolism	9.71	0.0182	0.0314	2	PLA2G4E, AGPAT1

CELLULAR PROCESSES					
Phagosome	7.62	0.0073	0.0207	3	TAP2, SEC61G, THBS1
ENVIRONMENTAL INFORMATION PROCESSING					
MAPK signaling pathway	5.80	0.0052	0.0207	4	FGFR2, RASGRP1, STK4, PLA2G4E
ECM-receptor interaction	9.14	0.0204	0.0314	2	THBS1, TNXB
ORGANISMAL SYSTEMS					
Fat digestion and absorption	16.89	0.0063	0.0207	2	PLA2G4E, AGPAT1
Salivary secretion	8.73	0.0222	0.0314	2	DMBT1, BST1
Fc gamma R-mediated phagocytosis	8.27	0.0246	0.0322	2	PLA2G4D, PLA2G4E
Pancreatic secretion	7.69	0.0281	0.0341	2	BST1, PLA2G4E
GENETIC INFORMATION PROCESSING					
Protein processing in endoplasmic reticulum	7.06	0.009	0.0217	3	ATF6B, SEC61G, RNF5
HUMAN DISEASES					
Hypertrophic cardiomyopathy (HCM)	9.36	0.0195	0.0314	2	SGCG, PRKAB2

Table 5: GO analysis using WebGestalt tool

GO Id	Description	Enrichment Ratio	rawP	adjP	No. of Genes	Genes
BIOLOGICAL PROCESS						
GO:0048568	embryonic organ development	5.89	8.31E-07	0.0009	12	*HEY2, HS6ST1, INSIG2, FGFR2, ALX4, HOXA2, HOXA5, SHH, TPO, HOXA6, SOX17, STK4*
GO:0003002	regionalization	5.79	2.95E-06	0.001	11	*CYP26B1, HEY2, FGFR2, ALX4, PRKDC, HOXA2, HOXA5, SHH, HOXA6, SOX17, SP8*
GO:0007389	pattern specification process	4.89	2.25E-06	0.001	13	*CYP26B1, HEY2, FOXK1, FGFR2, ALX4, PRKDC, HOXA2, HOXA5, SHH, HOXA6, SOX17, SP8, STK4*
GO:0009790	embryo development	3.25	7.56E-06	0.002	18	*CYP26B1, INSIG2, FOXK1, PRKDC, TPO, EXT2, SP8, HEY2, DMBT1, HS6ST1 ALX4, FGFR2, HOXA2, HOXA5, SHH, HOXA6, SOX17, STK4*
GO:0048705	skeletal system morphogenesis	7.27	1.42E-05	0.003	8	*CYP26B1, INSIG2, FGFR2, ALX4, HOXA2, HOXA5, HOXA6, THBS1*
GO:0060484	lung-associated mesenchyme development	48.77	2.62E-05	0.0046	3	*FGFR2, HOXA5, SHH*

GO Id	Description	Enrichment Ratio	rawP	adjP	No. of Genes	Genes
GO:0009952	anterior/posterior pattern specification	6.25	4.17E-05	0.0062	8	*HEY2, ALX4, PRKDC, HOXA2, HOXA5, SHH, HOXA6, SOX17*
GO:0048598	embryonic morphogenesis	3.94	4.90E-05	0.0064	12	*CYP26B1, INSIG2, FGFR2, ALX4, HOXA2, HOXA5, SHH, HOXA6, SOX17, EXT2, SP8, STK4*
GO:0048514	blood vessel morphogenesis	4.13	6.83E-05	0.0076	11	*HEY2, HS6ST1, FGFR2, HOXA5, SHH, ADAM8, ESM1, SOX17, XDH, STK4, THBS1*
GO:0043009	chordate embryonic development	3.78	7.30E-05	0.0076	12	*HEY2, HS6ST1, DMBT1, FGFR2, ALX4, PRKDC, HOXA2, HOXA5, SHH, HOXA6, SOX17, STK4*
GO:0009792	embryo development ending in birth or egg hatching	3.74	8.15E-05	0.0077	12	*HEY2, HS6ST1, DMBT1, FGFR2, ALX4, PRKDC, HOXA2, HOXA5, SHH, HOXA6, SOX17, STK4*
GO:0060523	prostate epithelial cord elongation	108.37	0.0001	0.008	2	*FGFR2, SHH*
GO:0051150	regulation of smooth muscle cell differentiation	32.51	9.71E-05	0.008	3	*HEY2, FGFR2, SHH*
GO:0060916	mesenchymal cell proliferation involved in lung development	81.28	0.0002	0.0123	2	*FGFR2, SHH*
GO:0001568	blood vessel development	3.63	0.0002	0.0123	11	*HEY2, HS6ST1, FGFR2, HOXA5, SHH, ADAM8, ESM1, SOX17, XDH, STK4, THBS1*
GO:0060737	prostate gland morphogenetic growth	81.28	0.0002	0.0123	2	*FGFR2, SHH*
GO:0061031	endodermal digestive tract morphogenesis	81.28	0.0002	0.0123	2	*FGFR2, SOX17*
GO:0001944	vasculature development	3.46	0.0003	0.0165	11	*HEY2, HS6ST1, FGFR2, HOXA5, SHH, ADAM8, ESM1, SOX17, XDH, STK4, THBS1*
GO:0001525	angiogenesis	4.1	0.0003	0.0165	9	*HS6ST1, FGFR2, HOXA5, SHH, ADAM8, ESM1, SOX17, STK4, THBS1*
GO:0048562	embryonic organ morphogenesis	5.24	0.0004	0.0199	7	*INSIG2, HOXA6, FGFR2, ALX4, HOXA2, HOXA5, SHH*
GO:0045165	cell fate commitment	5.1	0.0004	0.0199	7	*CYP26B1, HEY2, FGFR2, SOX17, PRKDC, HOXA2, SHH*
GO:0060441	epithelial tube branching involved in lung morphogenesis	19.51	0.0005	0.0237	3	*FGFR2, HOXA5, SHH*
GO:0048565	digestive tract development	7.39	0.0006	0.0241	5	*FGFR2, SOX17, ALX4, HOXA5, SHH*
GO:0001501	skeletal system development	3.78	0.0006	0.0241	9	*CYP26B1, INSIG2, FGFR2, ALX4, HOXA2, HOXA5, SHH,HOXA6, THBS1*
GO:0048706	embryonic skeletal system development	7.46	0.0006	0.0241	5	*HOXA6, ALX4, HOXA2, HOXA5, SHH*
GO:0030855	epithelial cell differentiation	4.24	0.0006	0.0241	8	*HEY2, DMBT1, FGFR2, HOXA5, SHH, SOX17, STK4, XDH*
GO:0010467	gene expression	1.51	0.0008	0.0246	44	*BANP, SIGIRR, ATOH8, INSIG2, LIME1, SHQ1, ATF6B, SEC61G, SLC2A4RG,*

GO Id	Description	Enrichment Ratio	rawP	adjP	No. of Genes	Genes
						KLF13, TRIM27, EXT2, SP8, THBS1, HEY2, IL16, PADI4, GALNT9, H2AFY, PIWIL3, CRY2, SOX17, MRPL15, CEBPD, RYBP, CYP26B1, FOXK1, ZNF500 PRKDC, NAA38, ST3GAL5, SP4, CUZD1, MAN2A2, HS6ST1, FGFR2, ALX4, HOXA2, HOXA5, SHH, NCOA7, HOXA6, ADAM8, XDH
GO:0060462	lung lobe development	46.44	0.0008	0.0246	2	*FGFR2, SHH*
GO:0060463	lung lobe morphogenesis	46.44	0.0008	0.0246	2	*FGFR2, SHH*
GO:0034766	negative regulation of ion transmembrane transport	46.44	0.0008	0.0246	2	*TRIM27, THBS1*
GO:0060664	epithelial cell proliferation involved in salivary gland morphogenesis	46.44	0.0008	0.0246	2	*FGFR2, SHH*
GO:0060349	bone morphogenesis	9.7	0.0008	0.0246	4	*CYP26B1, INSIG2, FGFR2, THBS1*
GO:0055123	digestive system development	6.89	0.0008	0.0246	5	*FGFR2, SOX17, ALX4, HOXA5, SHH*
GO:0001570	vasculogenesis	9.56	0.0008	0.0246	4	*HEY2, SOX17, XDH, SHH*
GO:0060439	trachea morphogenesis	40.64	0.001	0.0268	2	*HOXA5,SHH*
GO:2001212	regulation of vasculogenesis	40.64	0.001	0.0268	2	*HEY2, XDH*
GO:0051151	negative regulation of smooth muscle cell differentiation	40.64	0.001	0.0268	2	*HEY2, SHH*
GO:0033089	positive regulation of T cell differentiation in thymus	40.64	0.001	0.0268	2	*ADAM8, SHH*
GO:0045647	negative regulation of erythrocyte differentiation	40.64	0.001	0.0268	2	*KLF13, HOXA5*
GO:0035108	limb morphogenesis	6.25	0.0012	0.0291	5	*CYP26B1, FGFR2, ALX4, SP8, SHH*
CELLULAR COMPONENT						
GO:0031225	anchored to membrane	6.64	0.0003	0.0158	6	*GGT1, DPEP1, ALPPL2, ALPP, BST1, GFRA4*
GO:0012505	endomembrane system	2.16	0.0002	0.0158	24	*CYP26B1, MAP1LC3C, INSIG2, ATF6B, SEC61G, AGPAT1, SLC35C1, TRIM27, ST3GAL5, EXT2, CUZD1, PLA2G4D, MAN2A2, DMBT1, HS6ST1, GALNT9, FITM2, DYSF, TAP2, C8orf17, LRRC59, ADAM8, RNF5, RASGRP1*
GO:0042589	zymogen granule membrane	39.84	0.0011	0.0385	2	*CUZD1, DMBT1*

DISCUSSION

This is the first large-scale epigenome-wide site-specific collaborative meta-analysis studying the relationship between plasma Hcy levels and methylation at a genome wide level. Our results show that plasma Hcy concentrations were associated with 3 DMPs and 68 DMRs. Identification of this relatively low number of DMPs is surprising, given the central role of Hcy in the methylation cycle.

The most significant of the 3 DMPs, cg21607669 is located at chromosome 19 near the promoter region of the gene *SLC27A1* (+40 bp downstream). *SLC27A1* assists in the transport of fatty acids across cell membrane, and its highest levels are found in muscle and adipose tissue [62]. Results from a recent meta-analysis demonstrated that omega-3 polyunsaturated fatty acid supplementation was associated with Hcy lowering [63]. It was hypothesized that the omega-3 fatty acid, docosahexaenoic acid (DHA) upregulates metabolic enzymes of the one-carbon pathway like methionine adenosyl-transferase resulting in increased SAM bioavailability [63]. The exact role of hypermethylation of SLC27A1 should be further investigated to understand its possible role in the relation between polyunsaturated fatty acids and HHcy.

The second significant DMP, cg26382848 is located at chromosome 14 near the promoter region of the gene *AJUBA* (-39 bp upstream). *AJUBA* plays a role in cellular processes such as cell migration, proliferation and differentiation [64, 65]. It functions as a negative regulator of the retinoic signalling [66] and Hippo signalling pathway [67], and is involved in tumors [68]. However, the precise role of this gene is less clear and future replication studies are necessary to establish the role of this gene in relation to Hcy.

The third significant DMP cg10701000 at chromosome 10 annotated to the nearby gene *KCNMA1* (-143,626 bp upstream). *KCNMA1* is a potassium calcium-activated channel and has its prominent role in many physiological processes of smooth muscles, hair cells and diseases of nervous system and cancer [69]. A recent study has shown that elevated cellular Hcy increases the activity of potassium calcium-activated channel in GH3 pituitary cells [70], while two other previous studies have reported that Hcy inhibits these channels of artery smooth muscle cells [71, 72]. Our finding of hypermethylation near the KCNMA1 gene supports the involvement of Hcy in potassium calcium-activated channel.

We specifically studied the IGF2/H19 locus, the most frequently studied locus in relation to Hcy levels, and found that, none of the CpGs on the array that were annotated to this locus showed an association with Hcy. We, could therefore not corroborate results from 2 earlier studies that did find a relationship between Hcy levels and methylation at the IGF2/H19 locus [14, 15]. The reason could be that these studies were performed in HHcy mice models with heterogeneous knockout of the *CBS* gene, fed with high methionine and low folate diet. CBS deficiency is an inborn error of metabolism which leads to Hcy concentrations much higher at intermediate to severe

levels, as compared to the ones present in our study. Furthermore, in human studies of patients with HHcy and uremia, the influence of Hcy on H19 methylation is also observed at a concentration of Hcy much higher at intermediate levels, as compared to the mild levels present in our study, and the monoallelic expression of H19 is shown to be reversed by folate supplementation [10]. Therefore, our results cannot be directly extrapolated in other populations where Hcy is higher especially in chronic renal failure and rare diseases of metabolism. However, our results are in line with another human study which did not find differences in intermediate HHcy patients with CBS deficiency and controls [36]. These patients were under Hcy-lowering therapy which would have prevented the association with H19 DMR. Therefore, mildly elevated Hcy levels do not seem to be related to differences in methylation of the IGF2/H19 region in circulating leukocytes. Furthermore, focusing our analysis on the previously found CpGs in newborns in relation to maternal plasma folate during pregnancy, we did not find significant enrichment. We expected some enrichment because of the strong association of Hcy with folate. The reason for not finding enrichment could be related to the different study design of mother-offspring relationship compared to population-based cohorts.

Regional analysis using the comb-p software identified 68 DMRs significant at Sidak P less than 0.05. The most significant was 1.8 Kb DMR located at chromosome 6 spanning 55 CpGs. This DMR was annotated to the gene *TNXB* and *ATF6B*. *TNXB* is located within the class III region of the major histocompatibility complex [73] had multiple CpG sites previously shown to be hypermethylated in anorexia nervosa patients compared to controls [74, 75]. Another gene located near this DMR is the Activating Transcription Factor 6 Beta (ATF6B) gene. This gene is involved in the unfolded protein response during ER stress and has been shown to be activated upon Hcy treatment to human endothelial cells [76]. More specifically a putative ATF6-binding motif was identified, which was shown to be demethylated upon treatment with Hcy, supporting our findings of possible involvement of altered methylation of ATF6B in relation to HHcy.

Among the other DMRs there were a number of interesting regions that had prior studies connecting Hcy to the genes annotated near the CpGs. For example, the DMR Chr.6:32,145,146-32,146,780 is annotated to the genes *AGPAT1* and *RNP5*. This region was previously found associated to serum vitamin B12 in elderly subjects with elevated Hcy in our previous study of B-vitamins for the PRevention Of Osteoporotic Fractures (B-PROOF) [77]. Other DMRs, Chr.7:27,142,100-27,142,811 and Chr.7:27,187,269-27,187,692 were annotated to the HOX genes, which were previously shown to be differentially methylated after a 2-year intervention with folic acid and vitamin B12 [77]. Another gene *DPEP1* annotated to the DMR Chr.16:89,689,811-89,690,263, is involved in renal function and contains SNPs previously shown to be associated with Hcy [78-80]. Furthermore, gene PADI4 of DMR Chr.1:17,634,543-

17,634,717 previously showed to citrullinate DNMT3A, and therefore interacting with it to control DNA methylation [81]. And lastly, 2 out of 8 CpGs of DMR Chr.5:54,281,198-54,281,734 were significantly shown associated with maternal plasma folate levels during pregnancy, in cord blood [55]. These regions are promising to be replicated in future studies.

DMR analysis is recommended as a part of the analysis pipeline in addition to CpG site analysis [82, 83]. It has been shown that methylation is regulated in genomic regions, which merits the identification of DMRs in addition to single CpGs. Several software packages are available to perform DMR analysis, with their own strengths and weaknesses (i.e. Bumphunter, DMRcate) [84]. In this paper we applied comb-p, which constructs regions based on the autocorrelation between the p-values at varying lags in a sliding window, which is a good approach for the unevenly spaced CpGs in the 450k array [56]. A recent paper has shown that Comb-p has good sensitivity. However, if the data signal is weak, comb-p has the tendency to return false-positives particularly in presence of low effect size [85]. We therefore decided to check the DMR results of comb-p in the Rotterdam Study subset of 700 individuals, and compare them with the DMR results of another package, DMRcate [84]. Of the 49 significant DMRs within the subset of 700 individuals we found using comb-p package, we saw an overlap of 60%, with the DMRs found using DMRcate (data not shown). Results of comp-b seem valid but further studies are also necessary to validate these findings further.

Pathway analysis on the 114 genes annotated to the 68 DMRs identified pathways related to metabolism such as folate biosynthesis, glycosaminoglycan biosynthesis, arachidonic acid metabolism and glycerophospholipid metabolism. This could suggest how elevated Hcy can affect other important pathways via DNA methylation impairment. Folate pathways contained 2 genes: ALPPL2 and ALPP. An important paralog of these genes, which is ALPL, previously showed polymorphisms that were associated with vitamin B6 that suggests its possible role in vitamin B6 catabolism [86]. However, these genes need to be further replicated because of its low number in the pathway. Results of the GO enrichment analysis showed many biological processes related to embryogenesis and development, where Hcy has most variedly shown to play a role by previous studies.

Most of the studies until now investigated this association by measuring DNA methylation at a global level. The strength of our study is that the association of mild HHcy with genome-wide DNA methylation using 450k arrays was studied. Second, this is the largest epigenome wide association study on Hcy including 2,035 individuals.

Limitations of our study are that clinical heterogeneity between cohorts may have prevented us of finding additional true significant associations. Cohorts were either population based, or cases with venous thromboembolism, or have a mildly increased risk to develop type 2 diabetes and/or cardiovascular disease. Hcy concentrations and related mechanisms may vary with such different

pathophysiologies [87]. Furthermore, Hcy median concentrations were generally low in each cohort, with only 10.2-32.3% of individuals in the mild HHcy range >15 µmol/L [Table 1], which could explain the subtle findings. Second, we acknowledge that Hcy levels were measured using different techniques in each study, either in fasting or non-fasting state, as well as there were differences in sample collections per cohort. However, the methylation analyses were done in each cohort separately, and the results of each separate study were then meta-analysed. Our results are therefore not biased by the differences in Hcy measurement methods, state of fasting, or sample collection. Nevertheless, these heterogeneity across cohorts could have resulted in null-findings. Third, we did not account for differences in nutritional folate and/or vitamin B12 intake and B-vitamin supplementation, or homocysteine-related diseases like renal insufficiency that could have confounded the results of our study [87]. We also did not account for the genetic background of individuals, especially MTHFR 677C>T that is associated with Hcy [88]. We could have accounted for these confounding factors, but not all of this data was available to us. In addition, these confounding factors like folate, vitamin B12 and MTHFR 677C>T belong to the same pathway and are highly correlated with Hcy. This would have cancelled out the effect of Hcy itself on DNA methylation. We had hypothesized that Hcy is associated with DNA methylation regardless of the cause. However, adjusting for MTHFR 677C>T in the linear mixed model analysis in RS-III and CODAM did not change the findings (data not shown). Forth, we did not account for population stratification in the methylation analysis of each cohort because all individuals were of European ancestry. But in order to check the influence of this, we reran the analysis only in the Rotterdam Study by additionally adjusting for 4 genetic PCs. The results still showed no significant DMPs (FDR<0.05). In addition, the 3 significant DMPs of meta-analysis showed same direction of effects with and without the 4 genetic PCs correction [Figure S6]. Therefore, adding additional genetic PCs as covariates did not seem necessary and moreover, would reduce statistical power. Fifth, Hcy effect could be exclusive to certain other tissues like heart, liver and brain. Future studies still need to explore such findings. And lastly, even though we included 2,035 individuals from all the possible studies to our knowledge who had Hcy and DNA methylation data measured at the same time point, statistical power might still be low to identify significant DMP signals. As such, replication was not easily tractable, a main limitation of our results.

CONCLUSIONS

In conclusion, our meta-analysis showed 3 DMPs at chromosome 19, 14 and 10 with nearby genes of *SLC27A1* and *AJUBA*, and 68 DMRs associated with Hcy concentrations. Knowing the prominent role of Hcy in donation of methyl groups, we expected to find more DMPs. However, DMR analysis does show promising findings, but their role in relation to one-carbon metabolism needs to be further investigated.

ADDITIONAL INFORMATION

Supplementary Information accompanies this paper at *https://www.futuremedicine.com*.

REFERENCES

1. van Meurs JB, Dhonukshe-Rutten RA, Pluijm SM, van der Klift M, de Jonge R, Lindemans J, et al. Homocysteine levels and the risk of osteoporotic fracture. N Engl J Med 2004;350:2033-41.
2. Boushey CJ, Beresford SA, Omenn GS, Motulsky AG. A quantitative assessment of plasma homocysteine as a risk factor for vascular disease. Probable benefits of increasing folic acid intakes. Jama 1995;274:1049-57.
3. Toole JF, Malinow MR, Chambless LE, Spence JD, Pettigrew LC, Howard VJ, et al. Lowering homocysteine in patients with ischemic stroke to prevent recurrent stroke, myocardial infarction, and death: the Vitamin Intervention for Stroke Prevention (VISP) randomized controlled trial. JAMA 2004;291:565-75.
4. Lonn E, Yusuf S, Arnold MJ, Sheridan P, Pogue J, Micks M, et al. Homocysteine lowering with folic acid and B vitamins in vascular disease. N Engl J Med 2006;354:1567-77.
5. Bonaa KH, Njolstad I, Ueland PM, Schirmer H, Tverdal A, Steigen T, et al. Homocysteine lowering and cardiovascular events after acute myocardial infarction. N Engl J Med 2006;354:1578-88.
6. van Wijngaarden JP, Swart KM, Enneman AW, Dhonukshe-Rutten RA, van Dijk SC, Ham AC, et al. Effect of daily vitamin B-12 and folic acid supplementation on fracture incidence in elderly individuals with an elevated plasma homocysteine concentration: B-PROOF, a randomized controlled trial. Am J Clin Nutr 2014;100:1578-86.
7. Fu W, Dudman NP, Perry MA, Young K, Wang XL. Interrelations between plasma homocysteine and intracellular S-adenosylhomocysteine. Biochem Biophys Res Commun 2000;271:47-53.
8. James SJ, Melnyk S, Pogribna M, Pogribny IP, Caudill MA. Elevation in S-adenosylhomocysteine and DNA hypomethylation: potential epigenetic mechanism for homocysteine-related pathology. J Nutr 2002;132:2361S-2366S.
9. Yi P, Melnyk S, Pogribna M, Pogribny IP, Hine RJ, James SJ. Increase in plasma homocysteine associated with parallel increases in plasma S-adenosylhomocysteine and lymphocyte DNA hypomethylation. J Biol Chem 2000;275:29318-23.
10. Ingrosso D, Cimmino A, Perna AF, Masella L, De Santo NG, De Bonis ML, et al. Folate treatment and unbalanced methylation and changes of allelic expression induced by hyperhomocysteinaemia in patients with uraemia. Lancet 2003;361:1693-9.
11. Kang SS, Wong PW, Malinow MR. Hyperhomocyst(e)inemia as a risk factor for occlusive vascular disease. Annu Rev Nutr 1992;12:279-98.
12. Weiss N, Keller C, Hoffmann U, Loscalzo J. Endothelial dysfunction and atherothrombosis in mild hyperhomocysteinemia. Vasc Med 2002;7:227-39.
13. Mandaviya PR, Stolk L, Heil SG. Homocysteine and DNA methylation: a review of animal and human literature. Mol Genet Metab 2014;113:243-52.
14. Devlin AM, Bottiglieri T, Domann FE, Lentz SR. Tissue-specific changes in H19 methylation and expression in mice with hyperhomocysteinemia. J Biol Chem 2005;280:25506-11.
15. Glier MB, Ngai YF, Sulistyoningrum DC, Aleliunas RE, Bottiglieri T, Devlin AM. Tissue-specific relationship of S-adenosylhomocysteine with allele-specific H19/Igf2 methylation and imprinting in mice with hyperhomocysteinemia. Epigenetics 2013;8:44-53.

16. Ma S, Zhang H, Sun W, Gong H, Wang Y, Ma C, et al. Hyperhomocysteinemia induces cardiac injury by up-regulation of p53-dependent Noxa and Bax expression through the p53 DNA methylation in ApoE(-/-) mice. Acta Biochim Biophys Sin (Shanghai) 2013;45:391-400.
17. Zhang D, Wen X, Wu W, Xu E, Zhang Y, Cui W. Homocysteine-related hTERT DNA demethylation contributes to shortened leukocyte telomere length in atherosclerosis. Atherosclerosis 2013;231:173-9.
18. Jadavji NM, Deng L, Leclerc D, Malysheva O, Bedell BJ, Caudill MA, Rozen R. Severe methylenetetrahydrofolate reductase deficiency in mice results in behavioral anomalies with morphological and biochemical changes in hippocampus. Mol Genet Metab 2012;106:149-59.
19. Sontag E, Nunbhakdi-Craig V, Sontag JM, Diaz-Arrastia R, Ogris E, Dayal S, et al. Protein phosphatase 2A methyltransferase links homocysteine metabolism with tau and amyloid precursor protein regulation. J Neurosci 2007;27:2751-9.
20. Devlin AM, Singh R, Bottiglieri T, Innis SM, Green TJ. Hepatic acyl-coenzyme a:cholesterol acyltransferase-2 expression is decreased in mice with hyperhomocysteinemia. J Nutr 2010;140:231-7.
21. Sulistyoningrum DC, Singh R, Devlin AM. Epigenetic regulation of glucocorticoid receptor expression in aorta from mice with hyperhomocysteinemia. Epigenetics 2012;7:514-21.
22. Devlin AM, Singh R, Wade RE, Innis SM, Bottiglieri T, Lentz SR. Hypermethylation of Fads2 and altered hepatic fatty acid and phospholipid metabolism in mice with hyperhomocysteinemia. J Biol Chem 2007;282:37082-90.
23. Farkas SA, Bottiger AK, Isaksson HS, Finnell RH, Ren A, Nilsson TK. Epigenetic alterations in folate transport genes in placental tissue from fetuses with neural tube defects and in leukocytes from subjects with hyperhomocysteinemia. Epigenetics 2013;8:303-16.
24. Kim CS, Kim YR, Naqvi A, Kumar S, Hoffman TA, Jung SB, et al. Homocysteine promotes human endothelial cell dysfunction via site-specific epigenetic regulation of p66shc. Cardiovasc Res 2011;92:466-75.
25. Rodriguez-Esparragon F, Serna-Gomez JA, Hernandez-Velazquez E, Buset-Rios N, Hernandez-Trujillo Y, Garcia-Bello MA, Rodriguez-Perez JC. Homocysteinylated protein levels in internal mammary artery (IMA) fragments and its genotype-dependence. S-homocysteine-induced methylation modifications in IMA and aortic fragments. Mol Cell Biochem 2012;369:235-46.
26. Huang YS, Zhi YF, Wang SR. Hypermethylation of estrogen receptor-alpha gene in atheromatosis patients and its correlation with homocysteine. Pathophysiology 2009;16:259-65.
27. Al-Ghnaniem R, Peters J, Foresti R, Heaton N, Pufulete M. Methylation of estrogen receptor alpha and mutL homolog 1 in normal colonic mucosa: association with folate and vitamin B-12 status in subjects with and without colorectal neoplasia. Am J Clin Nutr 2007;86:1064-72.
28. Pirouzpanah S, Taleban FA, Atri M, Abadi AR, Mehdipour P. The effect of modifiable potentials on hypermethylation status of retinoic acid receptor-beta2 and estrogen receptor-alpha genes in primary breast cancer. Cancer Causes Control 2010;21:2101-11.
29. Vasavi M, Ponnala S, Gujjari K, Boddu P, Bharatula RS, Prasad R, et al. DNA methylation in esophageal diseases including cancer: special reference to hMLH1 gene promoter status. Tumori 2006;92:155-62.

30. Naushad SM, Reddy CA, Kumaraswami K, Divyya S, Kotamraju S, Gottumukkala SR, et al. Impact of hyperhomocysteinemia on breast cancer initiation and progression: epigenetic perspective. Cell Biochem Biophys 2014;68:397-406.
31. Jiang Y, Zhang H, Sun T, Wang J, Sun W, Gong H, et al. The comprehensive effects of hyperlipidemia and hyperhomocysteinemia on pathogenesis of atherosclerosis and DNA hypomethylation in ApoE-/- mice. Acta Biochim Biophys Sin (Shanghai) 2012;44:866-75.
32. Chen Z, Karaplis AC, Ackerman SL, Pogribny IP, Melnyk S, Lussier-Cacan S, et al. Mice deficient in methylenetetrahydrofolate reductase exhibit hyperhomocysteinemia and decreased methylation capacity, with neuropathology and aortic lipid deposition. Hum Mol Genet 2001;10:433-43.
33. Choumenkovitch SF, Selhub J, Bagley PJ, Maeda N, Nadeau MR, Smith DE, Choi SW. In the cystathionine beta-synthase knockout mouse, elevations in total plasma homocysteine increase tissue S-adenosylhomocysteine, but responses of S-adenosylmethionine and DNA methylation are tissue specific. J Nutr 2002;132:2157-60.
34. Castro R, Rivera I, Struys EA, Jansen EE, Ravasco P, Camilo ME, et al. Increased homocysteine and S-adenosylhomocysteine concentrations and DNA hypomethylation in vascular disease. Clin Chem 2003;49:1292-6.
35. Baccarelli A, Wright R, Bollati V, Litonjua A, Zanobetti A, Tarantini L, et al. Ischemic heart disease and stroke in relation to blood DNA methylation. Epidemiology 2010;21:819-28.
36. Heil SG, Riksen NP, Boers GH, Smulders Y, Blom HJ. DNA methylation status is not impaired in treated cystathionine beta-synthase (CBS) deficient patients. Mol Genet Metab 2007;91:55-60.
37. Hsu CY, Sun CY, Lee CC, Wu IW, Hsu HJ, Wu MS. Global DNA methylation not increased in chronic hemodialysis patients: a case-control study. Ren Fail 2012;34:1195-9.
38. Hofman A, Brusselle GG, Darwish Murad S, van Duijn CM, Franco OH, Goedegebure A, et al. The Rotterdam Study: 2016 objectives and design update. Eur J Epidemiol 2015;30:661-708.
39. Westendorp RG, van Heemst D, Rozing MP, Frolich M, Mooijaart SP, Blauw GJ, et al. Nonagenarian siblings and their offspring display lower risk of mortality and morbidity than sporadic nonagenarians: The Leiden Longevity Study. J Am Geriatr Soc 2009;57:1634-7.
40. van Greevenbroek MM, Jacobs M, van der Kallen CJ, Blaak EE, Jansen EH, Schalkwijk CG, et al. Human plasma complement C3 is independently associated with coronary heart disease, but only in heavy smokers (the CODAM study). Int J Cardiol 2012;154:158-62.
41. Antoni G, Morange PE, Luo Y, Saut N, Burgos G, Heath S, et al. A multi-stage multi-design strategy provides strong evidence that the BAI3 locus is associated with early-onset venous thromboembolism. J Thromb Haemost 2010;8:2671-9.
42. Leitsalu L, Haller T, Esko T, Tammesoo ML, Alavere H, Snieder H, et al. Cohort Profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. Int J Epidemiol 2015;44:1137-47.
43. Graham Upton IC. Understanding Statistics. Oxford University Press 1996.
44. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, et al. High density DNA methylation array with single CpG site resolution. Genomics 2011;98:288-95.
45. Pidsley R, CC YW, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. BMC Genomics 2013;14:293.
46. Maksimovic J, Gordon L, Oshlack A. SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. Genome Biol 2012;13:R44.

47. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics 2014;30:1363-9.
48. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010;26:2190-1.
49. Bonder MJ, Luijk R, Zhernakova DV, Moed M, Deelen P, Vermaat M, et al. Disease variants alter transcription factor levels and methylation of their binding sites. Nat Genet 2017;49:131-138.
50. Devlin B, Roeder K. Genomic control for association studies. Biometrics 1999;55:997-1004.
51. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, et al. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol 2010;28:495-501.
52. Speir ML, Zweig AS, Rosenbloom KR, Raney BJ, Paten B, Nejad P, et al. The UCSC Genome Browser database: 2016 update. Nucleic Acids Res 2016;44:D717-25.
53. Koukoura O, Sifakis S, Soufla G, Zaravinos A, Apostolidou S, Jones A, et al. Loss of imprinting and aberrant methylation of IGF2 in placentas from pregnancies complicated with fetal growth restriction. Int J Mol Med 2011;28:481-7.
54. Guo L, Choufani S, Ferreira J, Smith A, Chitayat D, Shuman C, et al. Altered gene expression and methylation of the human chromosome 11 imprinted region in small for gestational age (SGA) placentae. Dev Biol 2008;320:79-91.
55. Joubert BR, den Dekker HT, Felix JF, Bohlin J, Ligthart S, Beckett E, et al. Maternal plasma folate impacts differential DNA methylation in an epigenome-wide meta-analysis of newborns. Nat Commun 2016;7:10577.
56. Pedersen BS, Schwartz DA, Yang IV, Kechris KJ. Comb-p: software for combining, analyzing, grouping and correcting spatially correlated P-values. Bioinformatics 2012;28:2986-8.
57. Dai H, Leeder JS, Cui Y. A modified generalized Fisher method for combining probabilities from dependent tests. Front Genet 2014;5:32.
58. Stouffer SA, ed et al. The American Soldier: Adjustment During Army Life, Volume I. Princeton University Press, Princeton, NJ 1949.
59. P L. On the Combination of Independent Tests. Magyar Tudomanyos. Akademia Matematikai Kutato Intezetenek Kozlemenyei, 3 1958:171–197.
60. Sida`k Z. Rectangular confidence region for the means of multivariate normal distributions. J. Am. Stat. Assoc. 1967:626–633.
61. Wang J, Duncan D, Shi Z, Zhang B. WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. Nucleic Acids Res 2013;41:W77-83.
62. Martin G, Nemoto M, Gelman L, Geffroy S, Najib J, Fruchart JC, et al. The human fatty acid transport protein-1 (SLC27A1; FATP-1) cDNA and gene: organization, chromosomal localization, and expression. Genomics 2000;66:296-304.
63. Dawson SL, Bowe SJ, Crowe TC. A combination of omega-3 fatty acids, folic acid and B-group vitamins is superior at lowering homocysteine than omega-3 alone: A meta-analysis. Nutr Res 2016;36:499-508.
64. Goyal RK, Lin P, Kanungo J, Payne AS, Muslin AJ, Longmore GD. Ajuba, a novel LIM protein, interacts with Grb2, augments mitogen-activated protein kinase activity in fibroblasts, and promotes meiotic maturation of Xenopus oocytes in a Grb2- and Ras-dependent manner. Mol Cell Biol 1999;19:4379-89.
65. Kanungo J, Pratt SJ, Marie H, Longmore GD. Ajuba, a cytosolic LIM protein, shuttles into the nucleus and affects embryonal cell proliferation and fate decisions. Mol Biol Cell 2000;11:3299-313.

66. Hou Z, Peng H, White DE, Negorev DG, Maul GG, Feng Y, et al. LIM protein Ajuba functions as a nuclear receptor corepressor and negatively regulates retinoic acid signaling. Proc Natl Acad Sci U S A 2010;107:2938-43.
67. Das Thakur M, Feng Y, Jagannathan R, Seppa MJ, Skeath JB, Longmore GD. Ajuba LIM proteins are negative regulators of the Hippo signaling pathway. Curr Biol 2010;20:657-62.
68. Tanaka I, Osada H, Fujii M, Fukatsu A, Hida T, Horio Y, et al. LIM-domain protein AJUBA suppresses malignant mesothelioma cell proliferation via Hippo signaling cascade. Oncogene 2015;34:73-83.
69. Contreras GF, Castillo K, Enrique N, Carrasquel-Ursulaez W, Castillo JP, Milesi V, et al. A BK (Slo1) channel journey from molecule to physiology. Channels (Austin) 2013;7:442-58.
70. Gaifullina AS, Yakovlev AV, Mustafina AN, Weiger TM, Hermann A, Sitdikova GF. Homocysteine augments BK channel activity and decreases exocytosis of secretory granules in rat GH3 cells. FEBS Lett 2016;590:3375-3384.
71. Au AL, Seto SW, Chan SW, Chan MS, Kwan YW. Modulation by homocysteine of the iberiotoxin-sensitive, Ca2+ -activated K+ channels of porcine coronary artery smooth muscle cells. Eur J Pharmacol 2006;546:109-19.
72. Cai B, Gong D, Pan Z, Liu Y, Qian H, Zhang Y, et al. Large-conductance Ca2+-activated K+ currents blocked and impaired by homocysteine in human and rat mesenteric artery smooth muscle cells. Life Sci 2007;80:2060-6.
73. Bristow J, Tee MK, Gitelman SE, Mellon SH, Miller WL. Tenascin-X: a novel extracellular matrix protein encoded by the human XB gene overlapping P450c21B. J Cell Biol 1993;122:265-78.
74. Booij L, Casey KF, Antunes JM, Szyf M, Joober R, Israel M, Steiger H. DNA methylation in individuals with anorexia nervosa and in matched normal-eater controls: A genome-wide study. Int J Eat Disord 2015;48:874-82.
75. Kesselmeier M, Putter C, Volckmar AL, Baurecht H, Grallert H, Illig T, et al. High-throughput DNA methylation analysis in anorexia nervosa confirms TNXB hypermethylation. World J Biol Psychiatry 2016:1-13.
76. Zhang D, Xie X, Chen Y, Hammock BD, Kong W, Zhu Y. Homocysteine upregulates soluble epoxide hydrolase in vascular endothelium in vitro and in vivo. Circ Res 2012;110:808-17.
77. Kok DE, Dhonukshe-Rutten RA, Lute C, Heil SG, Uitterlinden AG, van der Velde N, et al. The effects of long-term daily folic acid and vitamin B12 supplementation on genome-wide DNA methylation in elderly subjects. Clin Epigenetics 2015;7:121.
78. Pare G, Chasman DI, Parker AN, Zee RR, Malarstig A, Seedorf U, et al. Novel associations of CPS1, MUT, NOX4, and DPEP1 with plasma homocysteine in a healthy population: a genome-wide evaluation of 13 974 participants in the Women's Genome Health Study. Circ Cardiovasc Genet 2009;2:142-50.
79. van Meurs JB, Pare G, Schwartz SM, Hazra A, Tanaka T, Vermeulen SH, et al. Common genetic loci influencing plasma homocysteine concentrations and their effect on risk of coronary artery disease. Am J Clin Nutr 2013;98:668-76.
80. Zinck JW, de Groh M, MacFarlane AJ. Genetic modifiers of folate, vitamin B-12, and homocysteine status in a cross-sectional study of the Canadian population. Am J Clin Nutr 2015;101:1295-304.
81. Deplus R, Denis H, Putmans P, Calonne E, Fourrez M, Yamamoto K, et al. Citrullination of DNMT3A by PADI4 regulates its stability and controls DNA methylation. Nucleic Acids Res 2014;42:8285-96.

82. Wright ML, Dozmorov MG, Wolen AR, Jackson-Cook C, Starkweather AR, Lyon DE, York TP. Establishing an analytic pipeline for genome-wide DNA methylation. Clin Epigenetics 2016;8:45.
83. Bock C. Analysing and interpreting DNA methylation data. Nat Rev Genet 2012;13:705-19.
84. Peters TJ, Buckley MJ, Statham AL, Pidsley R, Samaras K, R VL, et al. De novo identification of differentially methylated regions in the human genome. Epigenetics Chromatin 2015;8:6.
85. Kolde R, Martens K, Lokk K, Laur S, Vilo J. seqlm: an MDL based method for identifying differentially methylated regions in high density methylation array data. Bioinformatics 2016;32:2604-10.
86. Tanaka T, Scheet P, Giusti B, Bandinelli S, Piras MG, Usala G, et al. Genome-wide association study of vitamin B6, vitamin B12, folate, and homocysteine blood concentrations. Am J Hum Genet 2009;84:477-82.
87. Selhub J. Homocysteine metabolism. Annu Rev Nutr 1999;19:217-46.
88. Murakami H, Iemitsu M, Sanada K, Gando Y, Ohmori Y, Kawakami R, et al. Associations among objectively measured physical activity, fasting plasma homocysteine concentration, and MTHFR C677T genotype. Eur J Appl Physiol 2011;111:2997-3005.

CHAPTER 5

Genetically defined elevated homocysteine levels do not result in widespread changes of DNA methylation in leukocytes

Pooja R. Mandaviya*, Roby Joehanes*, Dylan Aïssi*, Brigitte Kühnel*, Riccardo E. Marioni*, Vinh Truong*, Lisette Stolk, Marian Beekman, Marc Jan Bonder, Lude Franke, Christian Gieger, Tianxiao Huan, M. Arfan Ikram, Sonja Kunze, Liming Liang, Jan Lindemans, Chunyu Liu, Allan F. McRae, Michael M. Mendelson, Martina Müller-Nurasyid, Annette Peters, P. Eline Slagboom, John M. Starr, David-Alexandre Trégouët, André G. Uitterlinden, Marleen M.J. van Greevenbroek, Diana van Heemst, Maarten van Iterson, Philip S. Wells, Chen Yao, Ian J. Deary**, France Gagnon**, Bastiaan T. Heijmans**, Daniel Levy**, Pierre-Emmanuel Morange**, Melanie Waldenberger**, Sandra G. Heil**, Joyce B.J. van Meurs**, on behalf of the CHARGE Consortium Epigenetics group and BIOS Consortium

**These authors contributed equally as first authors*
***These authors contributed equally as senior authors*

PLoS One. 2017 Oct 30;12(10):e0182472.

ABSTRACT

Background: DNA methylation is affected by the activities of the key enzymes and intermediate metabolites of the one-carbon pathway, one of which involves homocysteine. We investigated the effect of the well-known genetic variant associated with mildly elevated homocysteine: *MTHFR* 677C>T independently and in combination with other homocysteine-associated variants, on genome-wide leukocyte DNA-methylation.

Methods: Methylation levels were assessed using Illumina 450k arrays on 9,894 individuals of European ancestry from 12 cohort studies. Linear-mixed-models were used to study the association of additive *MTHFR* 677C>T and genetic-risk score (GRS) based on 18 homocysteine-associated SNPs, with genome-wide methylation.

Results: Meta-analysis revealed that the *MTHFR* 677C>T variant was associated with 35 CpG sites in *cis*, and the GRS showed association with 113 CpG sites near the homocysteine-associated variants. Genome-wide analysis revealed that the MTHFR 677C>T variant was associated with 1 *trans*-CpG (nearest gene *ZNF184*), while the GRS model showed association with 5 significant *trans*-CpGs annotated to nearest genes *PTF1A*, *MRPL55*, *CTDSP2*, *CRYM* and *FKBP5*.

Conclusions: Our results do not show widespread changes in DNA-methylation across the genome, and therefore do not support the hypothesis that mildly elevated homocysteine is associated with widespread methylation changes in leukocytes.

INTRODUCTION

DNA methylation, an important epigenetic mechanism has gained interest in the field of cancer and aging over the last decade [1, 2]. DNA methylation is affected by the activities of the key enzymes and intermediate metabolites of the one-carbon pathway, one of which involves homocysteine (Hcy).

Our aim was to investigate the role of genetically defined Hcy levels on genome-wide DNA methylation. A number of earlier studies have reported a link between Hcy and DNA methylation [3]. In these studies, DNA methylation was quantified as a global measure, that represents the total methyl cytosine content of the DNA. In animal models, both diet- and genetically- induced elevated Hcy have been related to altered global methylation patterns in tissues of aorta, brain, liver and colon. In human subjects, global DNA methylation in blood was not consistently altered with elevated Hcy. The relationship between Hcy and methylation can be subject to substantial bias, given the strong relationship between several lifestyle factors, diseases and Hcy. A way to circumvent this bias is to use genetic factors determining Hcy concentrations as an instrument to study the relationship between Hcy and methylation. The use of genetically defined elevated Hcy eliminate the effects that are possibly caused by measurement errors, confounding and reverse causality. One of the most consistent genetic variants causing elevated Hcy is the *MTHFR* 677C>T (rs1801133), which explains 5.3% variance in Hcy [4]. Furthermore, we recently published 18 variants including *MTHFR* 677C>T to be robustly associated with Hcy [5]. The Genetic Risk Score (GRS) of these 18 Hcy-associated variants explained 5.9% variance in Hcy [5]. In the current study, we used *MTHFR* 677C>T independently and the combined weighted GRS of these 18 variants, to test whether genetically defined elevated Hcy concentrations are associated with DNA methylation changes in blood cells.

A number of studies [3] have examined the relationship between the *MTHFR* 677C>T variant and global DNA methylation in humans. In 5 studies, individuals with the *MTHFR* 677TT genotype were compared to those with the *MTHFR* 677CC genotype [6-10]. Lower global methylation in blood cells was observed in two studies [6, 7]. The remaining three studies showed no association in the lymphocyte or colonic tissue. All published studies until now had small sample sizes of less than 200. To the best of our knowledge, associations of genetically defined Hcy with site-specific CpG methylation on a genome-wide scale have not been done up to now. In order to investigate this, we analyzed DNA methylation data measured with the Infinium Illumina 450k arrays, in a large meta-analysis of 9,894 individuals comprising 12 cohorts. We hypothesize that genetically defined elevated Hcy is associated with altered DNA methylation.

MATERIALS AND METHODS

Study population

All participants provided a written informed consent, and each study was approved at the relevant organizations by their respective ethics review committees [RS, Institutional review board (Medical Ethics Committee) of the Erasmus Medical Center; LLS, Ethical committee of the Leiden University Medical Center; LL, Ethics committee of the University Medical Centre Groningen; NTR, Central Ethics Committee on Research Involving Human Subjects of the VU University Medical Centre; CODAM, Medical Ethical Committee of the Maastricht University; MARTHA, "Departement santé de la direction générale de la recherche et de l'innovation du ministère" (Projects DC: 2008-880 & 09.576); EGCUT; Research Ethics Committee of the University of Tartu; F5L, Research ethics boards of the University of Toronto and the Ottawa Hospital Research Institute; FHS, IRB (institutional review board); KORA, Local Ethics Committee; LBC1921, Lothian Research Ethics Committee (Wave 1: LREC/1998/4/183); LBC1936, Multi-Centre Research Ethics Committee for Scotland (Wave 1: MREC/01/0/56), and the Lothian Research Ethics Committee (Wave 1: LREC/2003/2/29)].

The analyses comprised of large population with 9,894 participants from 12 cohorts of European ancestry. Most of the cohorts were part of either the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium [11] and/or Biobank-based Integrative Omics Studies (BIOS) consortium [12]. All participants provided a written informed consent for the DNA collection and its use for genetic analyses. Each study was approved at the relevant organizations by their respective ethics review committees. Cohort-specific characteristics are provided in the S1 Text and S1 Table.

MTHFR 677C>T and homocysteine-associated SNPs

18 independent Hcy-associated SNPs from our GWAS meta-analysis [5], were selected to assess the relationship between mildly elevated Hcy concentrations and genome-wide DNA methylation. The genotypes of these SNPs were extracted from the genotyping data. Cohort-specific details of the quality control and the SNP imputation methods are provided in the S2 Table.

DNA methylation assessment

Whole blood samples were collected from the participants for DNA extraction. The genomic DNA was bisulfite converted using the Zymo EZ-96 DNA-methylation kit (Zymo Research, Irvine, CA, USA). Methylation profiling was performed using the Infinium Illumina HumanMethylation 450k BeadChip arrays (Illumina Inc., San Diego, USA) according to the manufacturers' protocol. Beta values from 0 to 1, which represent the percentage of methylation, were calculated from the extracted raw methylated (M) and unmethylated (U) probe intensities and a default alpha (α) of 100. This is defined

by the formula of β=M/(M+U+α). Normalization was performed on these raw beta values using DASEN [13] or SWAN [14] methods. Poor quality probes were excluded based on the detection p-values mostly >0.01 in >5% of samples. Cohort-specific data preprocessing methods are provided in the Table 1. Global methylation levels per sample was calculated by the mean of all CpGs as well as mean according to CpG islands, shores, shelves or non-coding regions [15].

Statistical analysis

Two models were run independently by each participating study. Firstly, an additive model for *MTHFR* 677C>T alone was used to investigate its independent association with genome-wide DNA methylation in a linear manner. For the *MTHFR* 677C>T variant, genotypes were coded as CC=0, CT=1 and TT=2 to study the effect in methylation per *MTHFR* 677T allele.

In the second analysis, a weighted Genetic Risk Scores (GRS) was constructed from all the 18 Hcy-associated variants to investigate their combined and additive effect on genome-wide DNA methylation. Weighted GRS were calculated on the basis of their effect sizes [5] and number of corresponding risk alleles. The product of the two was calculated for each SNP and then summed up for all SNPs. The GRS was calculated using the equation below, where N is the number of elevated Hcy causing risk alleles for each SNP (0, 1 or 2 per genotype).

$$\begin{aligned}\boldsymbol{GRS} = {} & 0.1583\ x\ N(rs1801133{:}A) + 0.0542\ x\ N(rs2275565{:}G) + \\ & 0.0718\ x\ N(rs234709{:}C) + 0.0435\ x\ N(rs4660306{:}T) + \\ & 0.0453\ x\ N(rs1801222{:}A) + 0.101\ x\ N(rs12134663{:}C) + \\ & 0.0529\ x\ N(rs12780845{:}A) + 0.056\ x\ N(rs2851391{:}T) + \\ & 0.0449\ x\ N(rs9369898{:}A) + 0.0422\ x\ N(rs838133{:}A) + \\ & 0.0864\ x\ N(rs7422339{:}A) + 0.1242\ x\ N(rs7130284{:}C) + \\ & 0.0963\ x\ N(rs154657{:}A) + 0.0597\ x\ N(rs548987{:}C) + \\ & 0.0395\ x\ N(rs42648{:}G) + 0.0512\ x\ N(rs2251468{:}C) + \\ & 0.045\ x\ N(rs957140{:}G) + 0.090\ x\ N(rs12921383{:}C) \qquad (1)\end{aligned}$$

Both the analyses were based on linear mixed models of lme4 package in R. We also analyzed the effect of *MTHFR* 677C>T and GRS on global methylation levels, where we calculated the overall mean levels per individual as well as categorized the means as per CGI annotations [15]. The models were adjusted for technical covariates and biological covariates like age, sex and differential white blood cell (WBC) counts (see S1 Table for details about covariates for each cohort). The technical covariates were cohort-specific and treated as random effects. WBC counts were either used as measured counts, or they were imputed based on the Houseman method as implemented in the minfi package [17], or the modified version of the Houseman method (Documentation and R script: https://github.com/mvaniterson/wbccPredictor)

that uses partial least-squares [18] to handle multivariate responses and high-dimensional covariates and has been previously used [19]. This method from van Iterson used the R package pls [18] to fit the linear model based on the DNA methylation data, to predict the white blood cell composition as percentages that sum up to almost 100%. Age and gender were used as covariates.

Meta-analysis

Summary statistics for the two models were obtained from each study. Because of the different probe exclusions in each cohort [Table1], we removed probes that were present in ≤4 studies. We also excluded probes with SNPs at single base extension site, and probes with improper binding [12], leaving a total of 465,694 probes for the meta-analysis. Meta-analysis was performed using the fixed effect model in METAL [20], with the classical approach that uses effect size estimates and standard errors as input obtained from the individual study summary statistics for each CpG probe. The output of meta-analysis gave the combined effect size estimates, standard errors and p-values per probe. These p-values were corrected using the Benjamini Hochberg method of false discovery rate (FDR), where FDR <0.05 was considered statistically significant. For the MTHFR 677C>T model, positive effect sizes correspond to percentage increase in methylation per *MTHFR* 677T allele. For the GRS model, positive effect sizes correspond to percentage increase in methylation per unit increase in GRS. We also took into account heterogeneity of the meta-analysis by I^2 which was calculated using METAL per probe and excluded significant probes if $I^2<40$. We also calculated the genomic inflation factor (λ) [21] to estimate the inflation in test statistics that may be caused by population structure or other unknown confounding factors. This λ was estimated for the distribution of p-values using the median method, which is defined as the ratio of the observed median of the test statistic distribution and the expected median 0.455 [21, 22].

Genomic Regions Enrichment of Annotations Tool (GREAT) was used for annotating CpGs for nearby genes, that assigns a basal regulatory region to extend up to 5 kb upstream and 1 kb downstream from its transcription start site and a maximum extension distance up to 1 Mb [23], as defined by UCSC [24]. Furthermore, strength of the instrument or allele score was calculated using the F-statistics, using the tool, mRnd [25]. We took into account cohort heterogeneity I^2 and excluded significant probes if $I^2<40$.

Identifying *cis*- and *trans*-CpG effects

We defined the CpGs as "*cis*" when the CpG was annotated within 1Mb upstream or downstream of the SNP. *Trans*-CpGs were defined as CpGs that were associated with the SNP, and were annotated >1Mb apart. We defined the CpGs in the GRS model the same way by accounting for the bp distance of the CpGs from each of the 18 SNPs. For

the significant *trans*-CpGs that were 1-5Mb apart from any of the 18 SNPs, we performed a conditional analysis adjusting for that SNP to investigate whether they were *trans*-CpGs associated with Hcy GRS or long range *cis*-CpGs driven by the nearby SNPs that were part of the GRS. In the conditional analysis, if the bonferroni corrected p-values were no longer significant, we considered those trans-CpGs as long-range cis-CpGs. For the significant *trans*-CpGs that were >5Mb apart from any of the 18 SNPs, we looked up for their tested individual association with each of the 18 SNPs in our previous *trans*-CpG mapping analysis [12]. This is to see whether the association of these *trans*-CpGs was Hcy GRS driven or driven by a single SNP that was a part of GRS. In order to confirm these *trans*-CpG effects, we performed a similar conditional analysis by including the respective SNP as a covariate in the model. Both conditional analyses were performed on a subset of 3,786 samples from 6 cohorts, and the results were compared with the unconditional analysis in the same subset.

H19/IGF locus

Three Differentially Methylated Regions (DMRs) of IGF2/H19 locus at chromosome 11 have been reported to be related with homocysteine [26, 27]. We identified seven CpGs on the 450k array that were underlying the three DMRs of this locus, for their association with MTHFR 677C>T variant or GRS. Bonferroni method was applied on these 7 CpGs to check for multiple testing.

Enrichment of folate-associated CpGs

We further focused our analysis on the 443 previously identified CPGs, of which methylation in cord blood of newborns were associated with maternal plasma folate levels [28]. We compared the p-values of these 443 CpGs from the MTHFR 677>T and the GRS results, and compared them to the p-values of 100 random CpGs with 1000 permutations, to check for their significant enrichment, using the Fisher's exact test.

Table 1. Details of methylation 450k pre-processing: Quality control , normalization and association model

	PROBES EXCLUSION				SAMPLE EXCLUSION	NORMALIZATION*		ASSOCIATION MODEL		
Cohorts	**Detection p-value criteria**	**Cross-reactive & polymorphic**	**XY**	**Final**	**Criteria (Method)**	**Dye bias correction**	**Normalization method**	**WBC counts**	**Technical covariates**	**Additional adjustments**
CHARGE CONSORTIUM										
RS-III	>0.01 in >5% samples	Included	Excluded	463,456	Sample Call Rate >99%, Poor bisulfite conversion, Failed chromosome X & Y clustering	No	DASEN	Measured	Array, array position	No
LBC1921	>0.01 in >5% samples	Included	Included	446,851	>0.01 det. p-value in >5% probes, Poor bisulfite conversion	No	None	Measured	Array, array position, plate, hybridization date	No
LBC1936	>0.01 in >5% samples	Included	Included	446,851	>0.01 det. p-value in >5% probes, Poor bisulfite conversion	No	None	Measured	Array, array position, plate, hybridization date	No
KORA	>0.01	Included	Included	441,487	>0.01 det. p-value in >20% probes	Yes	BMIQ	Imputed	Array, array position	No
FHS	None	Included	Included	485,512	Mismatched sex, Outliers based on principal components	No	DASEN	Imputed	Array, array position, principal components	No
BIOS CONSORTIUM										
RS	>0.01 in >5% samples	Included	Excluded	419,937	Poor bisulfite conversion	No	DASEN	Imputed	Array, array position	No
LLS	>0.01 in >5% samples	Included	Excluded	419,550	Poor bisulfite conversion	No	DASEN	Imputed	Array, array position	No
LLD	>0.01 in >5% samples	Included	Excluded	420,591	Poor bisulfite conversion	No	DASEN	Imputed	Array, array position	No
NTR	>0.01 in >5% samples	Included	Excluded	420,341	Poor bisulfite conversion	No	DASEN	Imputed	Array, array position	No
CODAM	>0.01 in >5% samples	Included	Excluded	410,042	Poor bisulfite conversion	No	DASEN	Imputed	Array, array position	No
OTHER COHORTS										
MARTHA	>0.05 in >5% samples	Excluded	Included	388,120	Sample PCA	Yes	SWAN	Measured	Array, array position	No
F5L	>0.05 in >5% samples	Excluded	Excluded	378,594	Sample PCA	Yes	SWAN	Imputed	Array, array position	Family structure

*NORMALIZATION: Background correction was done in all studies

META-ANALYSIS: Removed probes with SNPs at SBE & probes with improper binding [12], probes that were absent in ≥8 studies, *cis*-probes with <5 Mb distance from Hcy-SNPs

RESULTS

Population Characteristics

The meta-analysis included 9,894 adults from 12 cohorts. Studies were population-based, except for the Cohort on Diabetes and Atherosclerosis Maastricht, MARseille THrombosis Association study and the French-Canadian family study, where individuals were selected based on mildly increased diabetes mellitus type 2 and cardiovascular risk factors, cases of venous thrombosis and probands with venous thromboembolism, respectively.

Meta-analysis of MTHFR 677C>T and GRS model

The explained variance in Hcy by MTHFR 677C>T is 5.3% [4] and by GRS is 5.9% [5]. For a sample size of 9,894, the F-statistics of the additive MTHFR 677C>T and GRS was 554 and 621, respectively. Meta-analysis of 456,694 probes identified 35 *cis*- [S3 Table] and 1 *trans*- [S5 Table, Table 2] CpGs for the MTHFR model [Fig 1a] and 113 *cis*- [S4 Table] and 30 *trans*- [S6a, S7a and S8 Tables] CpGs for the GRS model [Fig 1b]. The λ was 1.01 for the MTHFR 677C>T SNP and 0.92 for GRS [S1 Fig].

***Cis*-CpGs**

Meta-analysis on 465,694 CpGs of the *MTHFR* 677C>T variant showed association with 35 *cis*- CpGs on chromosome 1 with FDR<0.05 [Fig 1a, S3 Table]. These *cis*-CpGs showed a range from 2.4% increase to 1.7% decrease in methylation per *MTHFR* 677T allele. The nearest genes associated with this *cis*-region included *MTHFR* itself, *AGTRAP*, *CLCN6*, *NPPA, NPPB, PLOD1, MFN2* and *TNFRSF8.* For the GRS model, we observed 113 *cis*-CpGs with FDR<0.05 [Fig 1b, S4 Table]. Out of the 113, 16 *cis*-CpGs showed overlap with the MTHFR 677C>T analysis, which involved a smaller region of 238 Kb [Fig 1c].

***Trans*-CpGs**

For the *MTHFR* 677C>T model, meta-analysis of 465,694 CpGs identified 1 significant *trans*-CpG which was located on chromosome 6 [Fig 1a, S5 Table]. This *trans*-CpG (cg05411165) showed 1% decrease in methylation per MTHFR T allele. It was annotated near *ZNF184* (25414 bp upstream) and *HIST1H2BL* (309398 bp downstream). For the GRS model, we observed 30 significant *trans*-CpGs [Fig 1b]. These *trans*-CpGs showed a range from 5.6% increase to 5.1% decrease in methylation per 0.1 unit increase in GRS. Of these 30 *trans*-CpGs, 23 were negatively associated with the GRS model. To assess overlap between two models, we evaluated association of the trans-CpG of the MTHFR 677 C>T model within the GRS model. This *trans*-CpG (cg05411165) showed a 10% decrease in methylation in the GRS model but was not FDR significant (raw p-value = 0.01).

Critical evaluation of the 30 *trans*-CpGs of GRS model demonstrated that 14 *trans*-CpGs were located in a large region of 3,08 Mb length within chromosome 6. The GRS model consists of 18 SNPs including a SNP on chromosome 6. The 14 trans CpG identified with the GRS model were at a distance between 1 and 5 Mb away from the Hcy-associated variant rs548987 of the *SLC17A3* gene at chromosome 6 [S6a Table].

Conditional analysis: Chromosome 6 region near rs548987

To investigate whether the 14 *trans*-CpGs on chromosome 6 near rs548987 were influenced by this variant, we performed conditional analysis on a subset of 3,786 samples from 6 cohorts. After correction of the model for rs548987 as a covariate, none of the 14 *trans*-CpGs were significant at a bonferroni threshold of 3.57E-03 [S6b Table].

Conditional analysis: Influence of SNPs within the GRS model

To further investigate the remaining 16 *trans*-CpGs from the 30, whether any of them were driven by a single variant, rather than the combined effect of the 18 homocysteine-associated variants, we checked the *trans*-CpG mapping analysis of the single SNPs using the BIOS dataset [12]. We observed that 7 of the 16 remaining *trans*-CpGs located >5 Mb from the Hcy-associated variants, were directly associated with either rs548987 SNP of *SLC17A3* gene at chromosome 6, or rs154657 SNP of *DPEP1* gene at chromosome 16 [S7a Table]. After correction for these 7 *trans*-CpGs by including the respective SNP as a covariate in the model, none of the 7 *trans*-CpGs remained significant at a bonferroni threshold of 7.14E-03 [S7b Table]. After correction for cis-effects of Hcy-associated SNPs in the GRS model, we identified a remaining list of 9 Hcy-associated *trans*-CpGs, 4 of which had substantial heterogeneity I^2 [S8 Table].

Overlapping *trans*-CpG between MTHFR and GRS models

When doing a lookup in the *MTHFR* 677C>T model for the finally identified 5 *trans*-CpGs, all of them showed similar direction of effect, but did not achieve genome-wide significance (Lowest raw p-value = 3.36E-03) [Table 2].

***Trans*-CpGs affecting Gene Expression**

We evaluated whether methylation levels of the observed *trans*-CpG from the *MTHFR* 677C>T model and 5 *trans*-CpGs from the GRS model were associated with expression levels of the nearby genes, in the BIOS dataset [12]. None of the *trans*-CpGs was associated with mRNA expression differences of nearby genes.

H19/IGF2 locus

We specifically focused on the IGF2-H19 region for differential methylation, since methylation at this locus has repeatedly been linked to the homocysteine metabolism in a number of studies [29-31]. S4 Fig shows the results of the whole IGF2-H19 region,

and the 7 CpGs annotated to 3 DMRs of the IGF2/H19 gene that had previously been reported to be differentially methylated (DMR0, DMR2, H19-DMR3). Data from our 450k arrays contained 2 CpGs at DMR0, 4 CpGs at DMR2 and 1 CpG at H19-DMR3. None of them showed an association with MTHFR 677C>T or GRS with a Bonferroni cut off of 7.14E-03 [S9 Table, S4 Fig].

Enrichment of previously found folate-associated CpGs

Next we focused on a set of 443 CpGs that were identified to be differentially methylated in children at birth according to the folate levels in the mothers [28]. We found a highly significant enrichment for significant p-values in the MTHFR model in the 443 CpGs as compared to a random set of other CpGs (>3 times enrichment of significant p-values, enrichment p=0.0079). However, we did not find a significant enrichment for the GRS model.

Global DNA methylation changes

In addition to genome-wide DNA methylation changes we analyzed the effect of *MTHFR* 677C>T and GRS models on overall mean methylation levels. There was no significant association between the *MTHFR* 677C>T or GRS on global methylation overall or mean methylation of CpG islands, shores, shelves or non-coding regions [Table 3] [15].

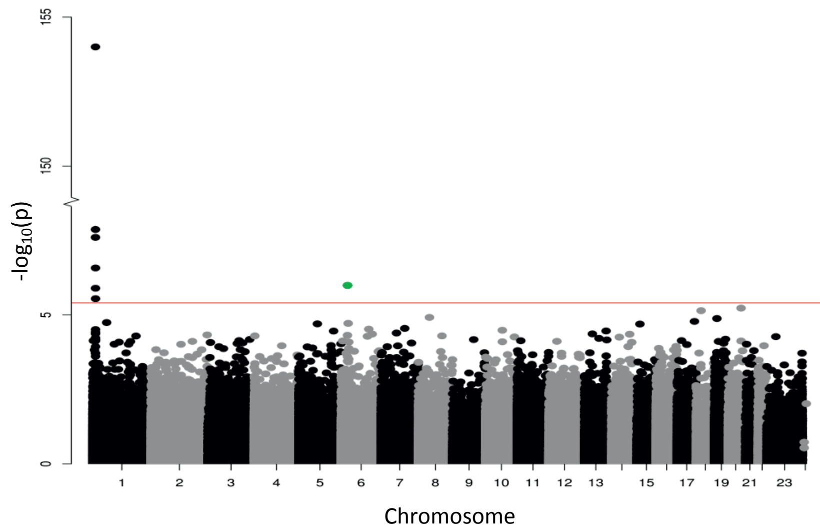


Figure 1a. Manhattan plot. Association between MTHFR 677C>T (rs1801133) and genome-wide DNA methylation in 9,894 samples, with 35 *cis*-meQTLs at chromosome 1 (black/grey) and 1 *trans*-meQTL at chromosome 6 (green) with FDR<0.05.

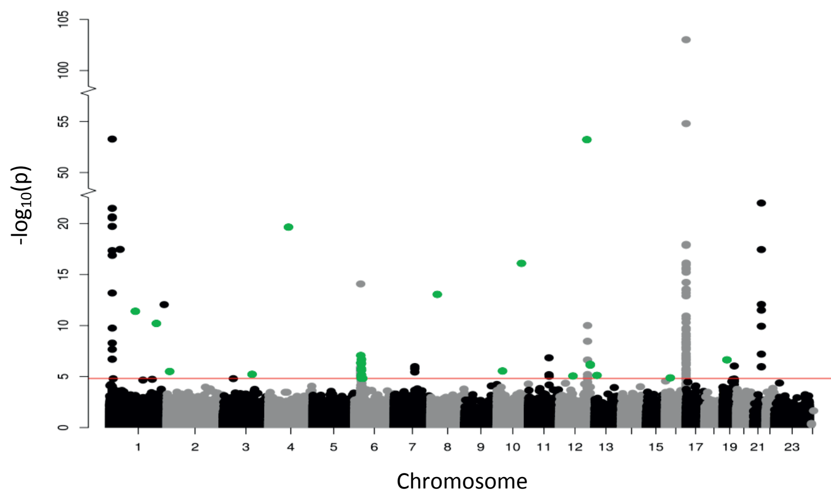


Figure 1b. Manhattan plot. Association between GRS of 18 Hcy-associated SNPs and genome-wide DNA methylation in 9,894 samples, with 113 *cis*-meQTLs (black/grey) and 30 *trans*-meQTLs (green), at FDR<0.05.

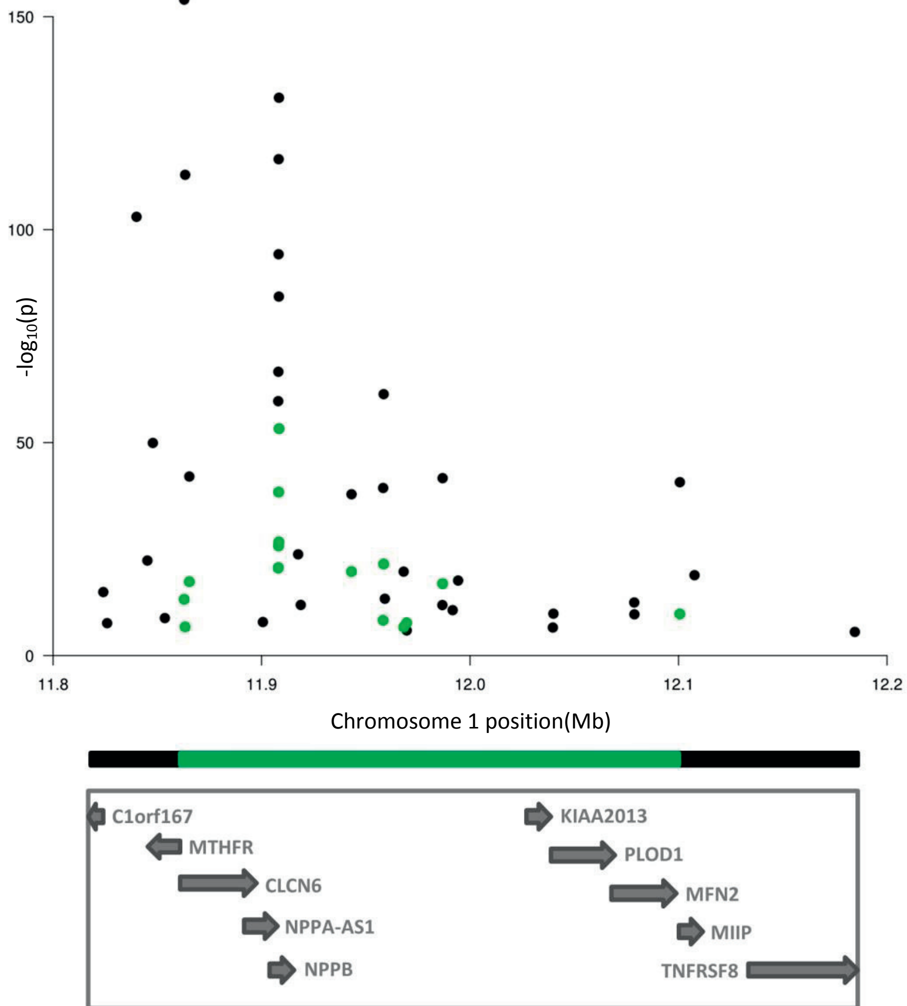


Figure 1c. Regional manhattan plot (chr1: 11824095- 12184574) for *MTHFR* 677C>T and GRS – cis overlap. 35 (black) and 16 (green) cis-meQTLs of the *MTHFR* 677C>T and GRS model respectively, in 9,894 samples. The overlap involved a small region of 238 kb (green rectangular line).

Table 2: Genome-wide *trans*-CpGs with FDR<0.05; associated with the *MTHFR* 677C>T model or Genetic Risk Score of 18 Hcy-associated variants.

CpG	N	Beta	SE	P	FDR	I^2	Beta	SE	P	Nearest Genes	Chr	Bp
		MTHFR 677C>T MODEL					LOOKUP IN GRS MODEL					
cg05411165	9894	-0.005	0.001	1.02E-06	1.40E-02	0	-0.0095	0.0038	1.16E-02	*ZNF184* (-25414), *HIST1H2BL* (+309398)	6	27466334
		GRS MODEL					LOOKUP IN *MTHFR* 677C>T MODEL					
cg12805629	6277	-0.018	0.004	3.15E-06	1.23E-02	0	-0.0023	0.0011	3.81E-02	*MRPL55* (+6698), *ARF1* (+19954)	2	11565653
cg08586216[+]	9894	-0.002	0.001	1.49E-05	4.89E-02	37.6	-0.0003	0.0001	4.17E-02	*TULP1* (-131681), *FKBP5* (+44391)	6	35612351
cg00620062	9894	0.004	0.001	2.83E-06	1.12E-02	0	0.0007	0.0003	3.36E-03	*PTF1A* (+6292)	10	23487775
cg00677455*	9334	-0.003	0.001	8.76E-06	3.09E-02	0	-0.0004	0.0002	3.63E-02	*CTDSP2* (-269)	12	58241039
cg01259782	6194	0.015	0.004	1.33E-05	4.52E-02	0	0.0021	0.0010	2.68E-02	*CRYM* (+454)	16	21313973

Beta: Regression coefficients, SE: Standard errors of the regression coefficients, FDR: False discovery rate adjusted P-value, threshold = 0.05, I^2: Heterogeneity I^2 parameter, *Promoter-associated, [+]Enhancer-associated, Enhancer annotation from Illumina 450k annotation

Table 3: Association of *MTHFR* 677C>T and Genetic Risk Score on mean global methylation levels

Methylation	N	I2	Beta	SE	P
MTHFR 677C>T					
GLOBAL	3,786	0.14	4.00E-06	1.50E-05	0.81
CGI	3,786	0.58	-1.90E-05	4.90E-05	0.70
SHE	3,786	0.59	3.70E-05	5.50E-05	0.50
SHO	3,786	0.14	-9.00E-06	4.80E-05	0.85
NC	3,786	0.61	2.70E-05	4.90E-05	0.57
GRS					
GLOBAL	3,786	0.67	-1.30E-05	5.60E-05	0.81
CGI	3,786	0.07	-1.84E-04	1.81E-04	0.31
SHE	3,786	0.00	2.29E-04	2.01E-04	0.25
SHO	3,786	0.52	-1.20E-04	1.76E-04	0.50
NC	3,786	0.07	1.94E-04	1.78E-04	0.28

Beta: Regression coefficients, SE: Standard errors of the regression coefficients, I^2: Heterogeneity I^2 parameter

CGI = CpG Islands, SHE = CpG Shelves, SHO = CpG Shores, NC = CpGs at Non-Coding regions

DISCUSSION

This is the first large-scale study to investigate the effect of Hcy-associated SNPs on genome-wide DNA methylation using the Illumina 450k arrays in 9,894 individuals. The results showed no widespread *trans*-effects of the *MTHFR* 677C>T SNP on DNA methylation, apart from 1 *trans*-CpG at chromosome 6. The GRS model showed 5 *trans*-CpGs, after carefully examining the direct effects of individual SNPs with conditional analyses.

In this current study, we used *MTHFR* 677C>T independently and the combined weighted GRS of the 18 Hcy-associated variants [5], to test whether mildly elevated Hcy concentrations induce DNA methylation changes in blood cells. Our goal was to investigate genetically defined elevated Hcy on genome-wide DNA methylation. The use of genetic variants is less sensitive to confounding and bias as compared to classical epidemiological studies [32].

We calculated the strength of our exposure variables: *MTHFR* 677C>T and GRS using the F-statistics. For a strong exposure, the value of the F-statistics is expected to be greater than 10 [33]. With our large sample size (n=9,894) and proportion of variance explained being 5.3 to 5.9%, the F-statistics was 554 and 621 respectively, indicating very high strength and enough power of our analysis. However, we did not observe widespread *trans*-effects despite of having strong additive *MTHFR* 677C>T and GRS.

We observed a single *trans*-CpG for the *MTHFR* 677C>T variant. This CpG is located near *ZNF184* and *HIST1H2BL*. Both these genes are thought to play a role in transcriptional regulation. For the GRS, we found 30 *trans*-CpGs associated, 14 of which are spread over a region of 3,08 Mb at chromosome 6. These CpGs were annotated to genes that included *ZNF322* and *HIST1H2BJ*, *HLA-J*, *HLA-A*, *HLA-G*, but also the proximal region of *ZFP57* gene, which was previously identified as a folate-sensitive region in a genome-wide methylation study of 23 women [34]. However, when we performed a conditional analysis on these 14 CpGs in this region, by adjusting for the nearby variant rs548987 of the *SLC17A3* gene, the effect sizes significantly attenuated and the nominal p-values were no more significant. The results indicate that this region was influenced by the rs548987 SNP of *SLC17A3* gene and was not Hcy-associated.

We finally observed 5 *trans*-CpGs associated to genetically defined Hcy using the GRS, after carefully examining the direct effects of individual SNPs with conditional analyses and discarding CpGs that showed substantial cohort heterogeneity I^2. A total of 3 CpGs showed hypomethylation, one of which was annotated to the *FKBP5* gene. *FKBP5* encodes for the FK506-binding protein 51 (FKBP51) whose expression has recently been shown to decrease DNMT1 activity and thereby decreasing global methylation [35].

Furthermore, when looking at our methylation-expression results [12], none of the 5 *trans*-CpGs was associated with mRNA expression differences of nearby genes.

The possible explanation for these negative findings could be that these CpG sites might have an effect further away on trans-genes. Conversely, it has been shown that the methylation-expression correlation in cis are not best predicted using the CpG position alone, but by using specific chromatin marks [36]. Furthermore, it could also be that these correlations are specific to other tissues, but not in blood.

We observed that the IGF2-H19 locus did not show association with methylation according to genetically defined elevated homocysteine. This is in contrast to the previous findings in mice, where tissue-specific changes in H19 DMR methylation were found in liver, brain and aorta, and increased expression of H19 was found in aorta [26, 27]. Similar to what we found, the H19-DMR3 between CBS deficient patients and controls also did not show a significant difference, in a previous study [29]. Our results show that this imprinted locus is not deregulated by long-term genetically defined mildly elevated homocysteine. However, previously it was reported that MTHFR 677C>T variant shows changes in DNA methylation in peripheral blood mononuclear cells, only through an interaction with folate [7]. Hence, further studies are needed to study the effect of MTHFR 677C>T variant in the presence of blood folate levels.

We did not see widespread methylation changes associated to mildly elevated plasma Hcy concentrations. This result is not in line with a number of earlier reports, which have shown global methylation changes in association with the *MTHFR* 677C>T variant and Hcy concentrations [3]. Previous two studies on this topic have shown contradictory results. There has been reports that showed a lower circulating global methylation level in individuals with the *MTHFR* 677TT genotype [6]'[7]. However, there are also a few negative studies that showed no relation between the *MTHFR* 677TT genotype and global methylation levels [8-10]. All these studies had modest sample sizes (upto 300 individuals were studied), and measured methylation on a global level using the LINE-1 assay, which measures a repetitive sequence, of which the function is unknown. In contrast, we here studied a genome wide site-specific analysis focused on functional regions of the genome [15]. We here show convincing evidence that there is no association between the *MTHFR* 677C>T or GRS on overall methylation levels, nor is there a relationship between methylation of CpG islands, shores, shelves or non-coding regions separately, which supports the previous null associations.

Furthermore, in order to test for causal effect in a mendelian randomization study, an instrument, which is in our case MTHFR 677C>T or GRS, should satisfy the 3 basic assumptions [37, 38]. One, the instrument should be associated with the exposure, which is in our case Hcy. Two, the instrument should not affect the outcome, which is in our case DNA methylation, except through the exposure Hcy. Three, the instrument should not be associated with any confounder of the exposure-outcome

association. Although assumptions one and two are satisfied in our case [5], the GRS model might violate assumption three [37, 38].

The GRS model contains a few SNPs which are, in addition to the association with Hcy, also associated with other traits. For example, the variants near to HNF1A gene have been associated with a number of other traits [39-44]. This could also be the reason why the results of the MTHFR 677C>T and GRS models are quite different. Nevertheless, the MTHFR 677C>T variant explains most of the variation in Hcy, as compared to the other variants in the GRS model and is therefore a strong instrument to examine the effect of deregulation of the one-carbon metabolism on methylation.

The relationship between *MTHFR* 677C>T and DNA methylation is modified by folate levels. Only in individuals with low folate status, the effect of the *MTHFR* 677TT genotype is seen [7]. Unfortunately, we were unable to study this interaction, since folate levels were not available in our study. Another prerequisite to be able to perform MR is that the relationship between Hcy and methylation is known. The relationship between Hcy and DNA methylation is only known in studies until now where methylation is measured at a global level. Therefore, the estimation of the causal effect could not be done. Unfortunately, we also did not have Hcy data available in all cohorts of this study, and therefore were unable to perform a full mendelian randomization study. We rather focused on the association of genetically defined elevated Hcy levels with DNA methylation.

We did not find widespread differences in methylation related to genetically defined homocysteine levels. The association was not observed in global methylation levels nor in widespread CpGs including the previously known H19/IGF2 locus. There are a number of possible explanations for this finding. First, it is known that the relationship between *MTHFR* 677C>T and DNA methylation is modified by folate levels, as described above. The effect of *MTHFR* 677C>T is seen in individuals with low folate status [7], which could have masked possible relationships between the MHTFR variant and methylation. A second possible explanation for the relative low number of identified CpGs, is that we have studied the wrong tissue. Most methylation measures are conducted in blood leukocytes as this tissue is readily available. However, the causal effect of Hcy could be specific to other tissues like liver, heart and brain. Therefore, the possible effect of mildly elevated Hcy on such specific tissues cannot be excluded. Third, there is little variation in the one-carbon metabolism in the normal population. This metabolism is pivotal to cell survival and function and therefore tidily regulated. It could be that there is a correlation between homocysteine and more pronounced effects on methylation when homocysteine levels are more extreme.

CONCLUSIONS

We observed 1 *trans*-CpG (nearest genes *ZNF184* and *HIST1H2BL*) on chromosome 6 associated with the *MTHFR* 677C>T variant. The GRS model showed 5 significant *trans*-CpGs, which do not overlap with the MTHFR trans-CpG. In conclusion, our results do not show widespread statistically significant *trans*-effects of MTHFR and GRS models, and therefore do not support the hypothesis that genetically defined mildly elevated Hcy concentrations are associated with widespread methylation changes in leukocytes. More studies with measured Hcy concentrations are needed to confirm this.

S1 Table. Cohort characteristics [Additional supplementary material accompanies this paper at *http://journals.plos.org/plosone.*]

Study Cohorts	N	Women [%]	MTHFR genotype %CT	MTHFR genotype %TT	GRS2 [Mean(S.D.)]	Age [Mean(S.D.)]	Gran [Mean(S.D.)]			Lym [Mean(S.D.)]				Mono [Mean(S.D.)]	Study Design
CHARGE consortium															
RS-III-I	729	54.0	41.6 (303)	8.6 (63)	1.09 (0.19)	59.86 (8.16)		4.21 (1.51)			2.42 (0.67)			0.42 (0.42)	Prospective, Population-based
LBC1921	436	60.3	45.0 (196)	12.4 (24)	1.13 (0.19)	79.10 (0.60)	4.6 (1.34)	0.17 (0.17)	0.02 (0.04)		1.71 (1.12)			0.54 (0.18)	Prospective, population-based
LBC1936	905	49.4	43.2 (391)	11.4 (103)	1.17 (0.18)	69.60 (0.80)	4.5 (1.54)	0.17 (0.13)	0.04 (0.04)		1.90 (1.43)			0.53 (0.19)	Prospective, population-based
KORA	1799	51.1	45 (779)	10.6 (183)	1.11 (0.18)	60.98 (8.89)		0.33 (0.09)		0.10[W] (0.07)	0.16[X] (0.06)	0.02[Y] (0.02)	0.05[Z] (0.03)	0.12 (0.03)	Prospective, population-based
FHS	2408	54.3	58.9 (1206)	18.2 (306)	1.12 (0.17)	66.4 (8.93)		0.49 (0.09)		0.10[W] (0.06)	0.19[X] (0.06)	0.02[Y] (0.02)	0.04[Z] (0.03)	0.13 (0.03)	Prospective, population-based
BIOS consortium															
RS-II-3/ RS-III-2	722	57.8	43.2 (312)	12.3 (89)	1.15 (0.19)	67.60 (5.98)	48.32[N] (7.04)	2.74[E] (0.73)	0.68[B] (0.18)		39.66 (9.71)			7.19 (1.30)	Prospective, population-based
LLS	739	51.7	48.0 (355)	11.8 (87)	1.14 (0.18)	58.98 (6.70)	58.14[N] (7.93)	2.63[E] (1.17)	0.64[B] (0.25)		29.52 (6.18)			5.68 (1.16)	Prospective, family-based
LLD	744	57.8	44.2 (329)	9.9 (74)	1.13 (0.17)	45.56 (13.32)	54.15[N] (7.22)	2.93[E] (1.49)	0.54[B] (0.25)		33.47 (7.10)			8.14 (1.53)	Prospective, population-based
NTR	691	68.5	41.8 (289)	10.3 (71)	1.12 (0.18)	33.95 (12.11)	52.59[N] (7.66)	2.81[E] (1.21)	0.47[B] (0.41)		35.11 (7.81)			7.72 (2.08)	-
CODAM	161	45.3	43.5 (70)	12.4 (20)	1.14 (0.17)	65.40 (6.77)	40.67[N] (6.97)	2.80[E] (0.63)	0.75[B] (0.20)		51.11 (14.06)			7.95 (1.63)	Prospective, Observational
Other cohorts															
MARTHA	349	78.5	48.1 (168)	11.7 (41)	1.14 (0.18)	43.77 (14.09)		4.58 (0.38)			1.87 (0.62)			0.32 (0.13)	Retrospective
F5L	211	52.1	48.8 (103)	20.9 (44)	1.15 (0.18)	39.7 (16.80)		0.56 (0.13)		0.04[W] (0.04)	0.22[X] (0.08)	0.10[Y] (0.05)	0.10[Z] (0.03)	0.08 (0.03)	Pedigree-based

[N] Neutrophils, [E] Eosinophils, [B] Basophils (Imputed), [W] CD8T, [X] CD4T, [Y] NK, [Z] Bcell (Imputed); Pathophysiology = Mainly healthy (except CODAM = Mildly increased DM2/CVD risk factors, MARTHA = Patients with venous thromboembolism, F5L = Probands with VTE)

REFERENCES

1. Ehrlich M. DNA methylation in cancer: too much, but also too little. Oncogene 2002;21:5400-13.
2. Klutstein M, Nejman D, Greenfield R, Cedar H. DNA Methylation in Cancer and Aging. Cancer Res 2016;76:3446-50.
3. Mandaviya PR, Stolk L, Heil SG. Homocysteine and DNA methylation: a review of animal and human literature. Mol Genet Metab 2014;113:243-52.
4. Borges MC, Hartwig FP, Oliveira IO, Horta BL. Is there a causal role for homocysteine concentration in blood pressure? A Mendelian randomization study. Am J Clin Nutr 2016;103:39-49.
5. van Meurs JB, Pare G, Schwartz SM, Hazra A, Tanaka T, Vermeulen SH, et al. Common genetic loci influencing plasma homocysteine concentrations and their effect on risk of coronary artery disease. Am J Clin Nutr 2013;98:668-76.
6. Castro R, Rivera I, Ravasco P, Camilo ME, Jakobs C, Blom HJ, de Almeida IT. 5,10-methylenetetrahydrofolate reductase (MTHFR) 677C-->T and 1298A-->C mutations are associated with DNA hypomethylation. J Med Genet 2004;41:454-8.
7. Friso S, Choi SW, Girelli D, Mason JB, Dolnikowski GG, Bagley PJ, et al. A common mutation in the 5,10-methylenetetrahydrofolate reductase gene affects genomic DNA methylation through an interaction with folate status. Proc Natl Acad Sci U S A 2002;99:5606-11.
8. Narayanan S, McConnell J, Little J, Sharp L, Piyathilake CJ, Powers H, et al. Associations between two common variants C677T and A1298C in the methylenetetrahydrofolate reductase gene and measures of folate metabolism and DNA stability (strand breaks, misincorporated uracil, and DNA methylation status) in human lymphocytes in vivo. Cancer Epidemiol Biomarkers Prev 2004;13:1436-43.
9. Hanks J, Ayed I, Kukreja N, Rogers C, Harris J, Gheorghiu A, et al. The association between MTHFR 677C>T genotype and folate status and genomic and gene-specific DNA methylation in the colon of individuals without colorectal neoplasia. Am J Clin Nutr 2013;98:1564-74.
10. Pufulete M, Al-Ghnaniem R, Rennie JA, Appleby P, Harris N, Gout S, et al. Influence of folate status on genomic DNA methylation in colonic mucosa of subjects without colorectal adenoma or cancer. Br J Cancer 2005;92:838-42.
11. Psaty BM, O'Donnell CJ, Gudnason V, Lunetta KL, Folsom AR, Rotter JI, et al. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from 5 cohorts. Circ Cardiovasc Genet 2009;2:73-80.
12. Bonder MJ, Luijk R, Zhernakova DV, Moed M, Deelen P, Vermaat M, et al. Disease variants alter transcription factor levels and methylation of their binding sites. Nat Genet. 2017 Jan;49:131-138.
13. Pidsley R, CC YW, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. BMC Genomics 2013;14:293.
14. Maksimovic J, Gordon L, Oshlack A. SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. Genome Biol 2012;13:R44.
15. Slieker RC, Bos SD, Goeman JJ, Bovee JV, Talens RP, van der Breggen R, et al. Identification and systematic annotation of tissue-specific differentially methylated regions using the Illumina 450k array. Epigenetics Chromatin 2013;6:26.
16. Bonder MJ, Kasela S, Kals M, Tamm R, Lokk K, Barragan I, et al. Genetic and epigenetic regulation of gene expression in fetal and adult human livers. BMC Genomics 2014;15:860.

17. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics 2014;30:1363-9.
18. Mevik B.H WR. The pls Package: Principal Component and Partial Least Squares Regression in R. Journal of Statistical Software 2007;18.
19. Dekkers KF, van Iterson M, Slieker RC, Moed MH, Bonder MJ, van Galen M, et al. Blood lipids influence DNA methylation in circulating cells. Genome Biol 2016;17:138.
20. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010;26:2190-1.
21. Devlin B, Roeder K. Genomic control for association studies. Biometrics 1999;55:997-1004.
22. Tsepilov YA, Ried JS, Strauch K, Grallert H, van Duijn CM, Axenovich TI, Aulchenko YS. Development and application of genomic control methods for genome-wide association studies using non-additive models. PLoS One 2013;8:e81431.
23. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, et al. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol 2010;28:495-501.
24. Speir ML, Zweig AS, Rosenbloom KR, Raney BJ, Paten B, Nejad P, et al. The UCSC Genome Browser database: 2016 update. Nucleic Acids Res 2016;44:D717-25.
25. Brion MJ, Shakhbazov K, Visscher PM. Calculating statistical power in Mendelian randomization studies. Int J Epidemiol 2013;42:1497-501.
26. Devlin AM, Bottiglieri T, Domann FE, Lentz SR. Tissue-specific changes in H19 methylation and expression in mice with hyperhomocysteinemia. J Biol Chem 2005;280:25506-11.
27. Glier MB, Ngai YF, Sulistyoningrum DC, Aleliunas RE, Bottiglieri T, Devlin AM. Tissue-specific relationship of S-adenosylhomocysteine with allele-specific H19/Igf2 methylation and imprinting in mice with hyperhomocysteinemia. Epigenetics 2013;8:44-53.
28. Joubert BR, den Dekker HT, Felix JF, Bohlin J, Ligthart S, Beckett E, et al. Maternal plasma folate impacts differential DNA methylation in an epigenome-wide meta-analysis of newborns. Nat Commun 2016;7:10577.
29. Heil SG, Riksen NP, Boers GH, Smulders Y, Blom HJ. DNA methylation status is not impaired in treated cystathionine beta-synthase (CBS) deficient patients. Mol Genet Metab 2007;91:55-60.
30. Koukoura O, Sifakis S, Soufla G, Zaravinos A, Apostolidou S, Jones A, et al. Loss of imprinting and aberrant methylation of IGF2 in placentas from pregnancies complicated with fetal growth restriction. Int J Mol Med 2011;28:481-7.
31. Guo L, Choufani S, Ferreira J, Smith A, Chitayat D, Shuman C, et al. Altered gene expression and methylation of the human chromosome 11 imprinted region in small for gestational age (SGA) placentae. Dev Biol 2008;320:79-91.
32. Palmer TM, Lawlor DA, Harbord RM, Sheehan NA, Tobias JH, Timpson NJ, et al. Using multiple genetic variants as instrumental variables for modifiable risk factors. Stat Methods Med Res 2012;21:223-42.
33. Pierce BL, Ahsan H, Vanderweele TJ. Power and instrument strength requirements for Mendelian randomization studies using multiple genetic variants. Int J Epidemiol 2011;40:740-52.
34. Amarasekera M, Martino D, Ashley S, Harb H, Kesper D, Strickland D, et al. Genome-wide DNA methylation profiling identifies a folate-sensitive region of differential methylation upstream of ZFP57-imprinting regulator in humans. Faseb J 2014;28:4068-76.

35. Gassen NC, Fries GR, Zannas AS, Hartmann J, Zschocke J, Hafner K, et al. Chaperoning epigenetics: FKBP51 decreases the activity of DNMT1 and mediates epigenetic effects of the antidepressant paroxetine. Sci Signal 2015;8:ra119.
36. Wagner JR, Busche S, Ge B, Kwan T, Pastinen T, Blanchette M. The relationship between DNA methylation, genetic and expression inter-individual variation in untransformed human fibroblasts. Genome Biol 2014;15:R37.
37. Swanson SA, Hernan MA. Commentary: how to report instrumental variable analyses (suggestions welcome). Epidemiology 2013;24:370-4.
38. Martens EP, Pestman WR, de Boer A, Belitser SV, Klungel OH. Instrumental variables: application and limitations. Epidemiology 2006;17:260-7.
39. Ligthart S, de Vries PS, Uitterlinden AG, Hofman A, group CIw, Franco OH, et al. Pleiotropy among common genetic loci identified for cardiometabolic disorders and C-reactive protein. PLoS One 2015;10:e0118859.
40. Below JE, Gamazon ER, Morrison JV, Konkashbaev A, Pluzhnikov A, McKeigue PM, et al. Genome-wide association and meta-analysis in populations from Starr County, Texas, and Mexico City identify type 2 diabetes susceptibility loci and enrichment for expression quantitative trait loci in top signals. Diabetologia 2011;54:2047-55.
41. Bonnycastle LL, Willer CJ, Conneely KN, Jackson AU, Burrill CP, Watanabe RM, et al. Common variants in maturity-onset diabetes of the young genes contribute to risk of type 2 diabetes in Finns. Diabetes 2006;55:2534-40.
42. Lauc G, Essafi A, Huffman JE, Hayward C, Knezevic A, Kattla JJ, et al. Genomics meets glycomics-the first GWAS study of human N-Glycome identifies HNF1alpha as a master regulator of plasma protein fucosylation. PLoS Genet 2010;6:e1001256.
43. Hegele RA, Cao H, Harris SB, Hanley AJ, Zinman B, Connelly PW. The private hepatocyte nuclear factor-1alpha G319S variant is associated with plasma lipoprotein variation in Canadian Oji-Cree. Arterioscler Thromb Vasc Biol 2000;20:217-22.
44. Pierce BL, Ahsan H. Genome-wide "pleiotropy scan" identifies HNF1A region as a novel pancreatic cancer susceptibility locus. Cancer Res 2011;71:4352-8.

CHAPTER 6

Interaction between plasma homocysteine and the MTHFR c.677C>T polymorphism is associated with site-specific changes in DNA methylation in humans

Alexander J Nash[1*], Pooja R Mandaviya[2,3*], Marie-Joe Dib[5], André G. Uitterlinden[3], Joyce van Meurs[3], Sandra G Heil[2**], Toby Andrew[4**] and Kourosh R Ahmadi[5**]

**These authors contributed equally as first authors*
***These authors contributed equally as senior authors*

Submitted for publication

ABSTRACT

One-carbon metabolism provides a direct link between dietary folate/vitamin B_{12} exposure, the activity of the enzyme methylenetetrahydrofolate reductase (MTHFR) and epigenetic regulation of the genome via DNA methylation. Previously, Friso et al. (2002) showed that the common *c.677C>T* polymorphism in *MTHFR* influences global DNA methylation status through a direct interaction with folate status and (indirectly) with homocysteine (tHcy) levels. To build on this observation, we investigated whether interaction between mildly elevated plasma tHcy and the *c.677C>T* polymorphism is associated with site-specific changes in DNA methylation in humans. We used data on plasma tHcy levels, *c.677C>T* polymorphism, and site-specific DNA methylation levels for a total of 915 Caucasian females and 335 males from the TwinsUK (n=610) and Rotterdam study (n=670). We carried out methylome-wide association analyses in each cohort to model the interaction between levels of tHcy and *c.677C>T* genotypes on DNA methylation beta values. Our meta-analysis identified a total of 13 probes significantly associated with [*rs1801133* x tHcy] levels (FDR<0.05). The most significant associations were with a cluster of probes at the *AGTRAP-MTHFR-NPPA/B* gene locus on chromosome 1 (FDR=1.3E-04), with additional probes on chromosomes 2, 3, 4, 7, 12, 16 and 19. Our top 2 hits on chromosome 1 were functionally associated with variability in expression of the *TNFRSF8* gene/locus on this chromosome. This is the first study to provide a direct link between perturbations in one-carbon metabolism, through an interaction of tHcy and the activity of MTHFR enzyme on epigenetic regulation of the genome via DNA methylation.

SIGNIFICANCE STATEMENT

A previous study by Friso et al. (PNAS 2002) has shown that folate status is associated with reduced genomic DNA methylation through a direct interaction with the common c.*677C>T* polymorphism in the *MTHFR* gene. This has highlighted an important epigenetic mechanism that links dietary exposure to aberrant genome regulation. To build on this observation, we sought to provide a direct link between perturbations in folate mediated one-carbon metabolism and epigenetic regulation of the genome via site-specific DNA methylation through an interaction between total plasma homocysteine - an integrative biomarker of one-carbon metabolism - and the activity of MTHFR enzyme. We identify a number of sites/loci that are dynamically regulated by this interaction.

INTRODUCTION

One-carbon metabolism (1-CM), comprising of the folate cycle and methionine pathway, plays fundamental roles in a variety of essential biochemical processes pertinent to healthy development and ageing, most notably provision of methyl residues (**-CH3**) to DNA, proteins and hormones [1]. Perturbations in 1-CM function can occur either through low dietary exposure to 1-carbon donors, the most studied of which is folate, malabsorption of the cofactor vitamin B_{12}, and/or genetic variation within genes that play important functions in normal 1-CM, most notably the *MTHFR c.677C>T (SNP rs1801133)*. The consequences of perturbed 1-CM, irrespective of the cause, include reduced remethylation of homocysteine (tHcy) into methionine leading to elevated plasma tHcy levels, as well as a reduction in cellular methylation potential [2].

Mild elevations in plasma tHcy has been unequivocally associated with increased risk of a myriad of adverse outcomes in observational studies, including cardiovascular and cerebrovascular disease, as well as increased rate of cognitive impairment and vascular dementia [3-5]. Whilst there is strong interest and growing evidence linking 1-CM function to aberrant DNA methylation [6-9], there are currently no systematic studies investigating the combined contribution of 1-CM function and the activity of the enzyme methylenetetrahydrofolate reductase (MTHFR) to epigenetic regulation of the genome via DNA methylation. Notably, a previous study by Friso *et al.* (2002) has shown that the common *c.677C>T* polymorphism in *MTHFR* influences global DNA methylation status through a direct interaction with folate status [10]. The Friso *et al.* study very elegantly showed that whilst global DNA methylation was independently associated with folate status, tHcy levels, and the *MTHFR c.677C>T* polymorphism, the association was principally driven by the interaction between the TT genotypes in *MTHFR* and low folate status culminating in a 50% reduction in the level of genomic DNA methylation in this group. Importantly, homozygosity for the *c.677C>T* polymorphism, especially in combination with low folate status, is also known to predispose to high plasma levels of fasting tHcy [11].

To build on these important observations and to systematically assess the impact of this interaction on site-specific DNA methylation patterns across the methylome in humans, we used total plasma Hcy (tHcy), *MTHFR c.677C>T* polymorphism, and site-specific DNA methylation levels from the Illumina 450k array of 1,280 individuals from two cohorts; TwinsUK (n=610) and Rotterdam study (n=670) respectively. We decided to focus on tHcy, as opposed to folate status, for two reasons. First, inadequate data coverage for folate meant that we were able to carry out the desired MWAS analysis in only n=200 samples from TwinsUK; folate levels were not available in the RS samples with methylation data. Secondly, we were more interested in unravelling the relationship between *MTHFR c.677C>T,* site-specific DNA methylation and tHcy. As tHcy has been shown to be a powerful integrative biomarker

of 1-CM function in health and disease we focused our analyses on this biomarker. Our overarching hypothesis was that elevated tHcy, a direct proxy for low folate status, in subjects harbouring at least one copy of the *rs1801133 T* allele is significantly associated with decreased DNA methylation levels at specific sites in the methylome. To this end, we performed a meta-analysis of the two cohorts to identify differentially methylated positions (DMPs) associated with *MTHFR c.677C>T* x tHcy interaction. Through this study we examined an important paradigm and gained novel mechanistic insight into the epigenetic consequences of this gene-nutrient interaction.

MATERIAL AND METHODS

Study populations

A total of 610 female twins from the TwinsUK registry (Table 1) and 670 participants from the Rotterdam study cohort with data on age, sex, height/weight used to calculate BMI, total plasma tHcy concentrations, smoking status, cell counts and site-specific DNA methylation based on the 450k array were included in the study. All participants were of North European descent and between 18 and 87 years old. In the TwinsUK cohort, we excluded subjects from the study that were on (i) any long-term medication(s) and/or with self-reported history of chronic diseases including hypertension or on any B-vitamin dietary supplements in the three months before participation. The Rotterdam cohort consisted of a randomly selected sub-sample of the Rotterdam study cohort (supplementary notes). We also excluded samples in both cohorts that showed evidence of being potential outliers or showing residual methylation array batch effects using iterative robust regression methods (see Methods).

This study was conducted according to the guidelines laid down in the Declaration of Helsinki and all procedures involving human subjects/patients were approved by the St. Thomas' Hospital Research Ethics Committee (EC04/015 TwinsUK) and Institutional review board (Medical Ethics Committee) of the Erasmus Medical Centre. Written informed consent was obtained from all subjects.

GENOME-WIDE DNA METHYLATION ANALYSIS

Methylation array data preparation and quality control

Genomic DNA was isolated from whole blood using the NORGEN DNA Purification Kit in TwinsUK (Norger Biotek Corporation) according to the manufacturer's protocol or by standardized salting out methods in Rotterdam. Approximately 700ng of each DNA sample was bisulphite converted using the EZ-96 DNA Methylation Kit (Zymo Research) according to the supplier's protocol. Bisulphite-treated DNA was then quantified using the NanoDrop ND-100 (NanoDrop Technologies).

Methylation profiling was performed on the bisulphite-converted DNA samples using the Illumina Infinium HumanMethylation450 BeadChip. Briefly, 5 µl of each bisulphite-converted DNA sample was processed following the Illumina protocol. The HumanMethylation450 BeadChip targets 485,764 methylation sites across the genome and interrogates the methylation state of each site using two assay designs: the Infinium I assay which has two bead types per CpG site, one each for the methylated and unmethylated states; and the Infinium II assay which has one bead type and determines methylation state with a single base extension step. The BeadChips were scanned with the Illumina HiScan SQ scanner and the raw data was imported to the GenomeStudio v.2010.3 software with the methylation module 1.8.2 for extraction of the image intensities.

Signal intensities for the methylated and un-methylated states were then quantile normalized for each probe type separately and beta values were calculated using R. The methylation state of each site is represented by a beta value (β), which is the ratio of the intensity of the methylated bead type (M) to the combined locus intensity (M+U+α) and ranges from 0 to 1. Data normalization [12] was performed on raw β values using Bioconductor packages implemented in R [13].

MTHFR 677C>T polymorphism

Genotypes for the *MTHFR c.677C>T* single nucleotide polymorphism (SNP *rs1801133*) were extracted from existing whole-genome Illumina/HumanHap array data from each cohort and utilised for the study analyses.

Exploratory Analysis and Model Fitting

Following Friso *et al*., we first used genomic methylation summary data from both the TwinsUK and Rotterdam Study cohorts to assess the relationship between genome-wide methylation and plasma tHcy levels and to control for technical variation in order to select a best-fit model to implement as part of the methylome wide association study. To do this, we first calculated the median genomic methylation level for each sample based upon the Illumina 450k array data that passed quality control. Using univariate and multiple regression analyses, median genomic methylation levels were regressed upon technical covariates to control for plate layout and batch effects, measured blood cell composition and key demographic variables such as age, sex, smoking status and body mass index.

Removal of outliers

Samples with different collection dates between methylation and tHcy data and samples that showed evidence of residual methylation array batch effects were identified. Only samples with tHcy values within the 95% of the distribution (between 5-30µmol/l) were retained and included in the analysis. Additional sample outliers

were identified and removed (TwinsUK n=40 and Rotterdam n= 27) by regressing genomic median methylation levels upon total tHcy plasma using iterative robust regression methods implemented in Stata (*rreg* procedure) and R (*rlm* function).

STATISTICAL ANALYSIS

Methylome-wide association study (MWAS)

DNA methylation at each Illumina probe was assessed for association with tHcy stratified by the MTHFR *c.677C>T* polymorphism (with *rs1801133* genotypes coded 0 for *CC* and 1 for *CT/TT*). In TwinsUK, a mixed linear regression model was implemented in R using the *LME4* package with DNA methylation at each probe regressed upon age, batch, cell count data and position on the BeadChip as fixed effects, and zygosity and family identifier included as random effects for the Null model. A similar model was implemented in Rotterdam Study, but batch effects were used as random effects. The Full model was identical to the Null but included two additional variables of interest to model the interaction between tHcy and the *rs1801133* genotype. The interaction model for each methylation probe was parameterised to include two nested interaction terms with the first parameter (tHcy0) used to estimate the regression coefficient between probe methylation and tHcy for individuals with the *rs1801133* genotype *CC* and a second parameter (tHcy1) to model the strength of association between methylation and tHcy for individuals with the *CT/TT* genotypes. Both regression lines were conservatively assumed to share the same intercept term for the MWAS. The potential interaction at each probe was then assessed by comparison of the full- and nested sub-model using a likelihood ratio (LLR) test. The model p value generated to assess potential differential methylation at each probe was calculated using the likelihood ratio test statistic, which is asymptotically Chi-squared with k degrees of freedom, where k is the number of model parameters estimated. For this test, the p-value of the effect of tHcy stratified by *c.677C>T* was calculated from the Chi-square distribution with 2 degrees of freedom. Multiple test correction was performed using the *qvalue* package in R. CpG probes with a false discovery rate (FDR) < 5% were considered significant.

Meta-analysis of MWAS results

All MWAS probe p-values from each cohort were meta-analysed using Fisher's combined probability test. Only those probes in which the direction of effect for the probe methylation response to tHcy concentration was consistent between the two cohorts were reported in the final meta-analysis (referred to here as concordant probes). This gave rise to a total of 182,227 concordant probes with potentially consistent results. All probes declared to be differentially methylated were also explicitly tested for a significant interaction term in order to ensure results were driven

by the *interaction* between tHcy and the *MTHFR* polymorphism that compromises folate metabolism and not just by tHcy levels or the genotype in isolation. To obtain a p-value for the interaction between *c.677C>T* and tHcy, a Wald test was used to assess the difference in slope for the two regression lines fitted to the nested interaction model (i.e. as presented in a scatterplot of methylation vs tHcy with two fitted regression lines). Equivalently, a cross classification model was used, which includes one main effect for *c.677C>T*, a main effect for tHcy and an interaction term between the two main effects (*c.677C>T* x tHcy). A multiple test correction was performed as described above. Probes that were significant with a false discovery rate of 5% or less, nominally significant and showing similar direction of effect in both cohorts were declared to be significant in the meta-analysis.

In-silico functional annotation of MWAS results

To characterise the genomic locations and properties of each region encompassing the most significant probes from the MWAS, the locations and relationship to specific genomic features were annotated using the Illumina Infinium HumanMethylation450 BeadChip manifest file and Genomic Regions Enrichment of Annotations Tool (great.stanford.edu). Finally, we looked at the association between any significant probes from the meta-analysis and differential gene expression using data from the BIOS QTL browser [14, 15].

Differentially Methylated Region Analysis

Where statistically significant probes from the meta-analysis were located in close proximity to one another other, the region was further analysed. For each of the top probes, the regression beta coefficients were stratified by genotype and plotted. A lowess fitted trend line was overlaid to assess the region as a differentially methylated site in response to disrupted folate and methionine metabolism.

RESULTS

A total of 1,280 Caucasian individuals from 2 independent cohorts – TwinsUK and Rotterdam study – with existing data on plasma tHcy, genotype data on the *MTHFR c.677C>T* polymorphism, and site-specific DNA methylation levels from the Illumina 450k array were included in the present study. Briefly, the TwinsUK cohort (n=610) was composed of generally healthy women with age ranging from 26-81 years (mean age= 58). The Rotterdam cohort (n=670) was a population-based sample of men and women (54%) with a mean age of 60 years. Table 1 provides demographic, biochemical, and descriptive summaries for each cohort.

Table 1. Study population characteristics for TwinsUK and Rotterdam Study samples. Values are presented as medians and standard deviations (SD).

	TwinsUK	Rotterdam
N	610	670
Age (years)	58.2 (9.4)	59.8 (8.1)
BMI (kg/m^2)	25.7 (4.8)	27.4 (4.8)
Smoking, % Never	13.7	27.3
Sex, % female	100	54
Methylation Beta	0.68 (0.02)	0.52 (0.02)
Homocysteine (µmol/L)	12.3 (3.4)	12.2 (3.6)
CELL COUNTS		
Granulocytes	4.00 (1.4)	4.2 (1.5)
Lymphocytes	1.8 (0.6)	2.4 (0.7)
Monocytes	0.4 (0.2)	0.4 (0.4)
Neutrophils	3.8 (1.4)	-

Exploratory Analysis and Model Fitting results

We carried out mixed linear and robust iterative regression analyses to test for the interaction between total genomic methylation and tHcy stratified by *MTHFR c.677C>T* genotype in both TwinsUK and Rotterdam cohort. Our analysis showed that the observed association between total genomic methylation and tHcy is primarily driven by individuals with high tHcy values and carrying one or more copies of the *MTHFR c.677C>T* allele. Whilst the interaction term is only marginally significant (P=0.08, adjusted for age, smoking status and technical covariates), the result is consistent with Friso *et al*. (PNAS 2002) and the over-arching hypothesis for this study. Figure 1A illustrates that the negative relationship between genomic methylation and tHcy is only significant for individuals with the *CT/TT* genotype (p= 0.02) and not for those with the *CC* genotype (p=0.93) in individuals from TwinsUK. In the Rotterdam cohort, we failed to replicate this finding and the results indicate a null relationship between genomic methylation and tHcy levels.

Our exploratory analyses identified estimated cell count and technical factors such as batch effect and position on the BeadChip to be strongly associated with total genomic DNA methylation. These were all included as nuisance covariates in the final model used for the methylome-wide association study (MWAS) to control for technical variation. Of the phenotypic variables, age had the largest influence on genomic DNA methylation.

The final model selected to take forward for the MWAS included tHcy levels stratified by *MTHFR rs1801133* (coded *CC* vs *CT/TT*), observed blood cell counts, age, BMI, batch effect and BeadChip position as fixed effects and family identifier and zygosity as random effects (TwinsUK) or alternatively, gender as fixed effect and array/bead chip positions as random effects (Rotterdam Study).

Cohort-specific MWAS and meta-analyses

We carried out individual cohort MWAS followed by a meta-analysis of the results generated for a total of 1,280 individuals of European ancestry (**Table 2**). The analysis of the TwinsUK data identified a total of 127 CpGs associated with [tHcy x *c.677C>T*] as the predictor (FDR≤ 5%). Analysis of the Rotterdam study alone included no significant probes at FDR threshold ≤ 5% for these data. **Figures 2** presents the *QQ* and Manhattan plots for the TwinsUK and Rotterdam study methylome scan. Meta-analysis identified 13 significant hits (Table 2) with consistent direction of effects between both cohorts. These included sites on chromosome 1 (meta-analysis p=3.7E-09, *NPPA*), chromosome 2 (p=6.2E-07, *DYNC2LI1*/*PLEKHH2*), chromosome 3 (p=7.5E-06, *VEPH1*/*SHOX2*), chromosome 4 (p=2.3E-06, *PROM1*), chromosome 7 (p=9.5E-07, *GLI3*/*C7orf25*), chromosome 12 (p=4.8E-06, *ALG10*), and chromosome 19 (p=5.4E-06, *HIF3A*).

Female-only MWAS and meta-analysis

When meta-analysis was restricted to females only for both cohorts – based on the rational that the TwinsUK cohort was primarily composed of female subjects – the total number of significant probes dropped from 13 to 7 but otherwise remained the same (chromosome 1: cg06770735, cg06193043; chromosome 3: cg04897900; chromosome 4: cg17858192; chromosome 12: cg10588893; chromosome 17: cg12432807 and chromosome 19: cg00192046). This is consistent with our assumption that the epigenetic signals at these loci do not differ between women and men, with the reduced number of observed significant probes due to the reduced power for the female-only analysis (n=960).

Chromosome 1 Analysis

Since all 4 significant probes on chromosome 1 (**Figure 3A & B**) were within 119Kb of one another, with two probes 148bp apart (cg06193043 and cg24844545), the extent to which the rest of the surrounding probes showed coordinated changes in DNA methylation in response to increasing tHcy were investigated. For the probes within a 50kb region surrounding the index probe *cg24844545* (indicated by a dashed line in the figure), there appears to be correlated responses to tHcy, with the direction of the response dependent on genotype at *rs1801133* (see **Figure 3C**) and the response consistent between the two cohorts. The correlated response at this locus is likely to indicate a [tHcy x *c.C677T*] dependent, differentially methylated region.

In-silico functional annotation of MWAS results

Finally, we used the BIOS QTL [14] browser to gain functional insight into the 13 significant probes from the meta-analysis and assess their putative impact on differential gene expression. Of the 13 sites, our top 2 hits on chromosome 1 (cg06193043 and cg24844545) were shown to be functionally associated with

variability in expression of the *TNFRSF8* gene/locus on this chromosome although no obvious link can be drawn between aberrant methylation at these sites, differential gene expression of the *TNFRSF8* gene/locus and 1-CM function. None of the other sites were shown to be associated with *cis* or *trans* differential gene expression.

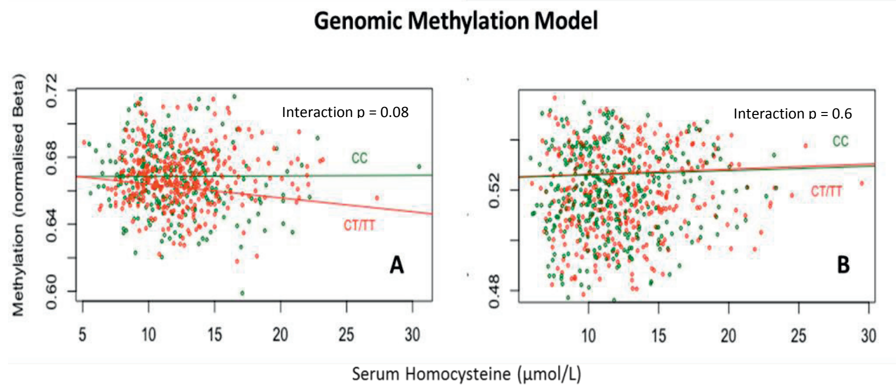
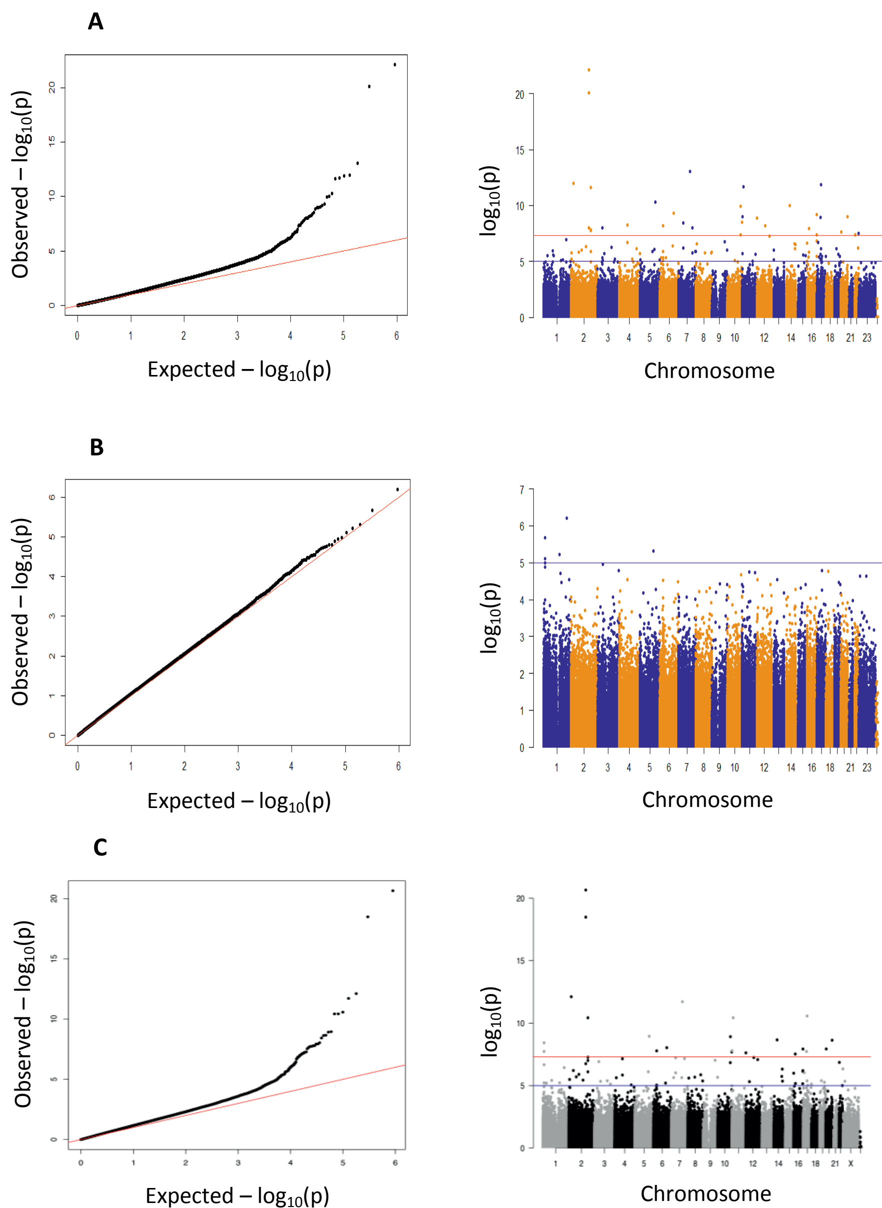


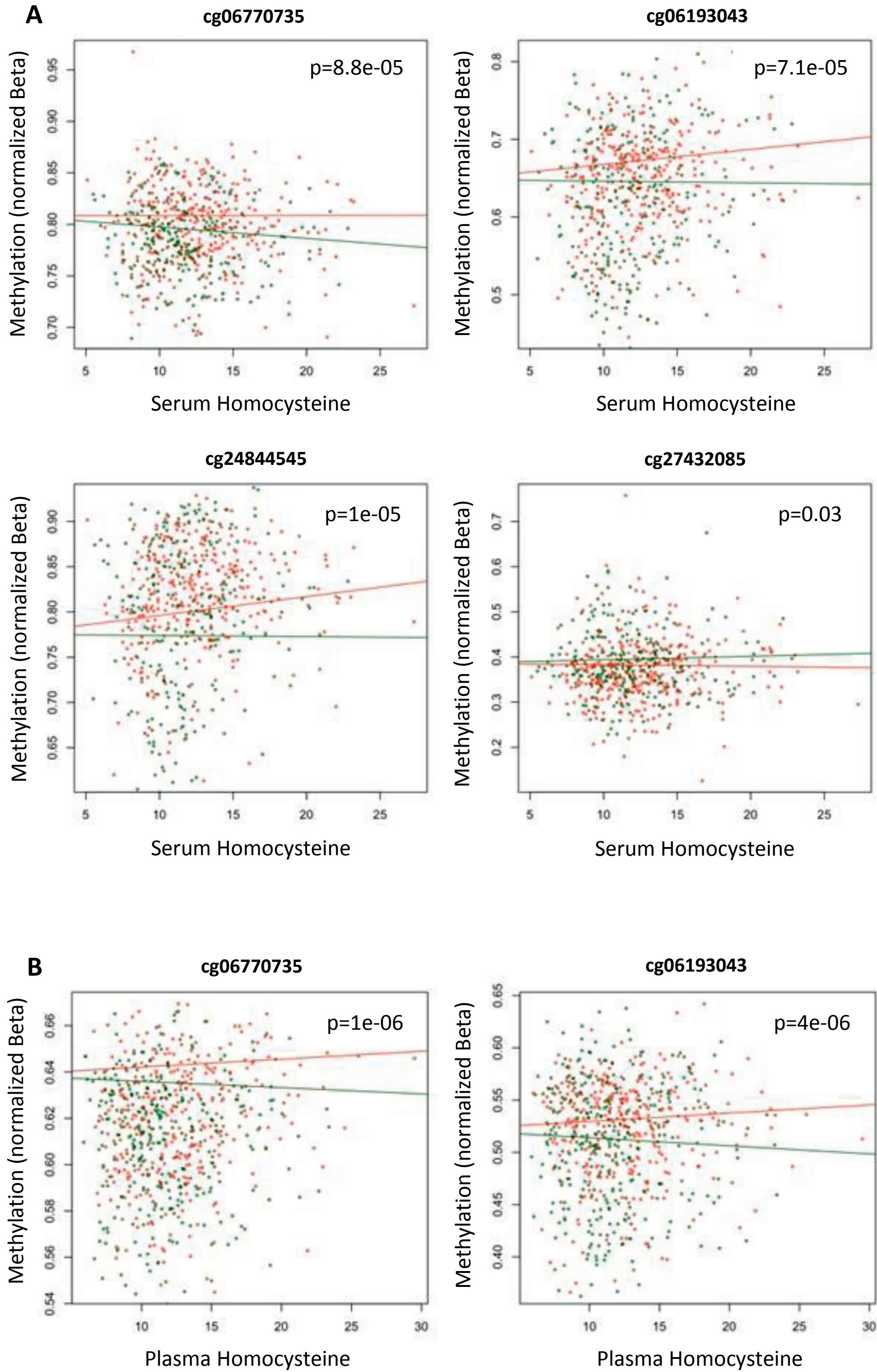
Figure 1. Graphical representation of the genomic methylation model. Median genomic methylation levels were regressed upon homocysteine (tHcy) stratified by *MTHFR c.677C>T* genotype, illustrating a differential impact of tHcy on genomic methylation levels by genotype in (A) TwinsUK and (B) Rotterdam. (A) TwinsUK data shows a significant negative relationship between median genomic methylation and serum homocysteine for individuals with rs1801133 genotypes CT/TT (p=0.02), but no relationship for individuals carrying the CC genotype (p=0.93; interaction or difference between the 2 slopes, p = 0.08). (B) Rotterdam data shows no significant relationship between median genomic methylation and serum homocysteine for all individuals, regardless of genotype.

Table 2. Genome-wide significant methylation probes identified from meta-analysis of TwinsUK (female) and Rotterdam Study (male & female) samples. Results are presented for probes with a significant FDR (q-value) of ≤ 0.05, nominally significant (p<0.10) and showing similar direction of effect in both cohorts.

Probe ID	CHR	Genomic position (kb)	Associated Genes	TwinsUK		Rotterdam		Meta-analysis	
				b	p-value	b	p-value	p-value	q-value
cg06770735	1	11,840,031	*MTHFR (+26104), AGTRAP (+43914)*	0.83	2.20E-04	1.33	3.90E-06	1.80E-08	3.60E-04
cg06193043	1	11,908,199	*NPPA (-384)*	1.23	1.30E-03	1.44	8.70E-06	2.10E-07	2.30E-03
cg24844545	1	11,908,347	*NPPA (-484)*	1.89	9.60E-05	1.38	1.70E-06	3.70E-09	1.20E-04
cg27432085	1	11,958,572	*NPPB (-39557), KIAA2013 (+27936)*	-0.23	9.50E-02	-1.81	4.00E-06	6.00E-06	3.20E-02
cg17110364	2	43,903,227	*DYNC2LI1 (-97974), PLEKHH2 (+38765)*	0.72	1.30E-04	2.6	2.60E-04	6.20E-07	5.70E-03
cg04706655	2	97,136,556	*NEURL3 (+34561), NCAPH(+135049)*	5.23	9.30E-07	0.83	7.80E-02	1.30E-06	9.90E-03
cg04897900	3	157,245,108	*VEPH1 (-23997), SHOX2 (+578819)*	1.29	6.00E-06	1.28	8.10E-02	7.50E-06	3.60E-02
cg17858192	4	16,077,807	*PROM1 (-43)*	0.98	1.40E-04	3	9.90E-04	2.30E-06	1.60E-02
cg24643514	7	23,221,283	*NUPL2 (-186)*	2.07	6.90E-02	5.24	2.00E-06	2.30E-06	1.50E-02
cg11119746	7	42,533,133	*GLI3 (-256540), C7orf25 (+418993)*	5.36	6.60E-07	2.15	8.10E-02	9.50E-07	7.80E-03
cg26161747	7	101,562,348	*SH2B2 (-366032), CUX1 (+101491)*	0.03	6.10E-03	3.78	2.20E-05	2.30E-06	1.50E-02
cg10588893	12	34,464,089	*ALG10 (+288898)*	3.82	3.60E-04	2.51	8.40E-04	4.80E-06	2.60E-02
cg02879662	19	46,800,467	*HIF3A (+189)*	2.23	1.80E-04	2.67	1.90E-03	5.40E-06	2.90E-02

Figure 2. QQ and Manhattan plots for the (A) TwinsUK, (B) Rotterdam Study full cohort MWASs, (C) and Meta-analysis.





A
cg06770735
p=8.8e-05
Methylation (normalized Beta)
Serum Homocysteine
cg06193043
p=7.1e-05
Methylation (normalized Beta)
Serum Homocysteine
cg24844545
p=1e-05
Methylation (normalized Beta)
Serum Homocysteine
cg27432085
p=0.03
Methylation (normalized Beta)
Serum Homocysteine
B
cg06770735
p=1e-06
Methylation (normalized Beta)
Plasma Homocysteine
cg06193043
p=4e-06
Methylation (normalized Beta)
Plasma Homocysteine

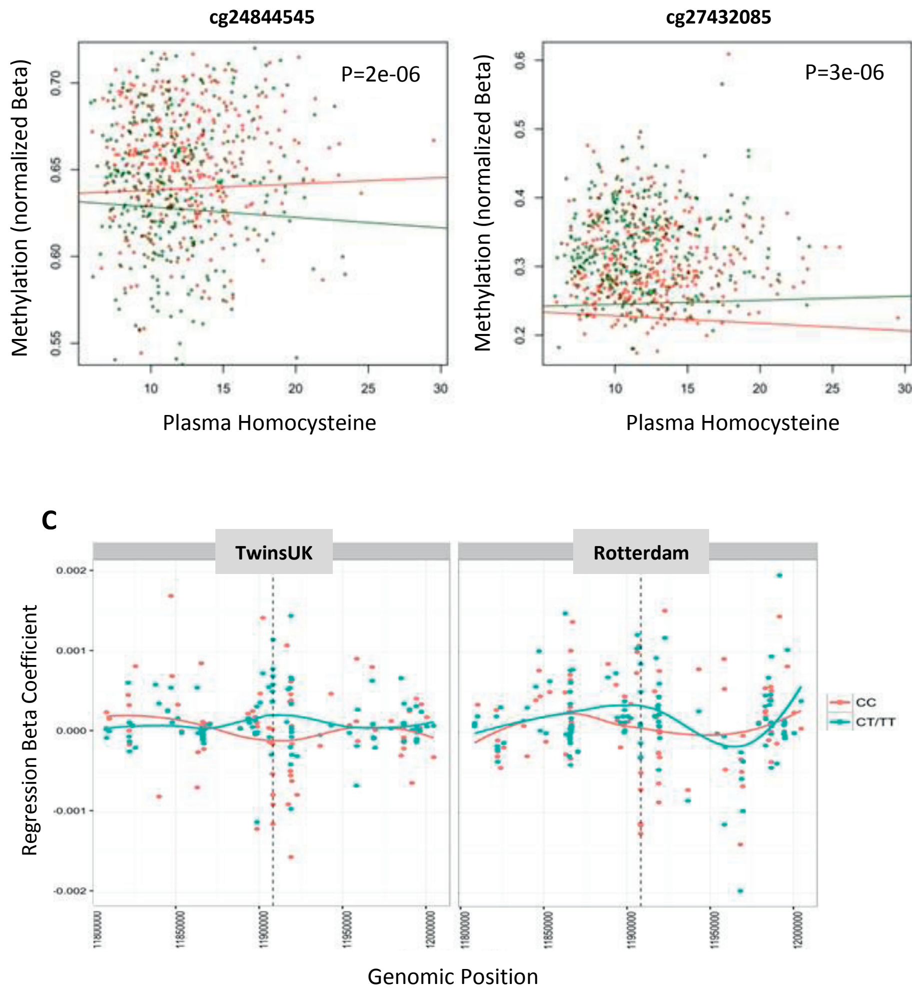


Figure 3. Chromosome 1 methylation probe association with tHcy is contingent upon *MTHFR* genotype: (A) TwinsUK and (B) Rotterdam Study. This figure illustrates the differential methylation for the probes on chromosome 1 with probe methylation plotted on the Y-axis and tHcy plotted on the X-axis with individuals distinguished by whether they carry a *CC MTHFR* c.*C677T* genotype (green) or a *CT/TT* genotype (red). All four chromosome 1 probes show an interaction between tHcy and SNP in both TwinsUK and Rotterdam Study. The strength of relationship between probe methylation and tHcy significantly differs depending upon the *MTHFR* genotype carried. (C) Correlated response to [tHcy x *c.C677T*] for the probes within a 50kb region surrounding the index probe *cg24844545* (indicated by a dashed line in the figure).

DISCUSSION

There is growing evidence suggesting that perturbations in 1-carbon metabolism (1-CM), in part due to variable dietary exposure or genetic variation - for example the common *MTHFR c.677C>T* polymorphism within the *MTHFR* gene - can affect patterns of DNA methylation and affect the regulation of gene expression throughout the course of life[2, 16]. A large number of epidemiological studies, including observational and experimental, have now shown associations between either low/normal levels of folate/vitamin B_{12} or mild elevations in plasma tHcy – an integrative marker and indicator of perturbed 1-CM - and increased risk of a range of vascular and neurodegenerative diseases[4, 17, 18]. Interestingly, a previous study by Friso et al. (2002) was the first to highlight that the relationship between folate status and global DNA methylation is strongly modulated by the *MTHFR c.677C>T* polymorphism and the activity of the MTHFR enzyme. However, to date there are no studies that have sought to identify differentially methylated sites in the human genome that are dynamically affected by the interaction between plasma tHcy and the *MTHFR c.677C>T* polymorphism.

To this end, we carried out a methylome-wide association study in two large, independent cohorts followed by a meta-analysis of the results, to identify specific sites across the methylome that are differentially methylated as a consequence of the interaction between *MTHFR c.677C>T* polymorphism and mild tHcy elevation.

Methylome-wide association studies of total homocysteine

Our initial exploratory analysis in TwinsUK underlined a significant negative relationship between DNA methylation and tHcy levels only with *CT/TT* genotype, but not *CC*. This is in accordance with previous studies that have investigated the interrelationships between concentrations of tHcy, the *MTHFR c.677C>T* polymorphism and global DNA methylation [19-21]. It's important to note that contrary to approach used by Friso et al.'s and subsequent studies [22, 23], our study did not explicitly use folate status and instead focused on tHcy as the principle exposure variable of interest. As explained previously in the introduction, the available TwinsUK data on folate was measured by a clinical laboratory that used a cut-off value of >20 ng/ml, all values above 20 ng/ml were assigned >20 ng/ml making nearly a third of the data unreliable and therefore not amenable to quantitative association analysis. We also had substantially fewer subjects with both folate and methylation data; N=200 as opposed to N= 610 with complete tHcy and methylation data. More importantly, however, although motivated by the findings from the Friso et al. study, we were more interested in unravelling the consequences of the interaction between aberrant 1-CM function, *MTHFR c.677C>T* and site-specific DNA methylation. Levels of total Hcy have been shown to be a powerful integrative biomarker of 1-CM function in health and disease and would allow a more holistic approach.

Our results show that the previous observations by Friso et al. (2002) on global DNA methylation are supported by our meta-analysis of 2 independent MWAS which identified a total of 13 differentially methylated positions (DMPs) on chromosomes 1, 2, 3, 4, 7, 12 and 19 significantly associated with [*c677C>T* x tHcy] levels. The most significant associations were with a cluster of probes at the *AGTRAP-MTHFR-NPPA/B* gene cluster on chromosome 1. *In-silico* analysis showed that differential methylation at this region is functionally associated with variability in expression of the TNF Receptor Superfamily Member 8 (*TNFRSF8)* gene[14].

Two recent MWASs by Mandaviya et al. (2017) have sought to identify associations between (1) tHcy; and (2) the *MTHFR C677T* variant and site-specific leukocyte DNA methylation levels [24, 25]. They identified 3 DMPs at or near genes *SLC27A1*, *AJUBA* and *KCNMA1* on chromosome 19, 14 and 10, respectively – as well as 68 differentially methylated regions (DMR) – associated with tHcy levels[24]. They further identified a DMR pertaining to the *MTHFR* locus in *cis* (≤ 1Mb upstream or downstream) and a single *trans*-CpG (> 1Mb upstream or downstream) associated with the *MTHFR C677T* variant [25]. Notably, the DMR at the *MTHFR* locus, identified by Mandaviya et al. to be associated with the *MTHFR c.677C>T* variant, overlaps with our top hits in the present study that include 4 *CpGs* at the *AGTRAP-MTHFR-NPPA/B* gene cluster on chromosome 1. So, could our top hit actually be a genetically-driven finding as opposed to a true consequence of the interaction between *MTHFR c677C>T* and tHcy given (1) the inherent lack of power associated with finding true interactions; and (2) that the *MTHFR* DMR was not associated with tHcy? The data provides strong evidence for an interaction between *rs1801133* genotype and serum homocysteine levels that impacts upon site-specific methylation levels and as such, we believe that our results represent a more realistic biological consequence of the interaction between 1-CM and genetic variation. It is also important to note that a true interaction between two or more exposure variables does not preclude main effect associations although a true biological interaction can manifest in the absence of independent main effects. So, it is very likely that both the *MTHFR c.677C>T* variant and the interaction [*c.677C>T* x tHcy] are associated with the DMR at the *MTHFR* locus. Furthermore, we would like to highlight that none of the other findings by Mandaviya et al. overlap with any of the remaining 9 loci identified at other chromosomes in the current study, emphasising the importance of appropriately modelling and identifying site-specific DNA methylation outcomes that are dynamically responsive to this interaction.

DMPs in AGTRAP-MTHFR-NPPA/B gene cluster on chromosome 1 and cardiovascular function

Our top hits pertain to a cluster of probes at the *AGTRAP-MTHFR-NPPA/B* gene locus on chromosome 1. The *NPPA* gene encodes for a protein from the natriuretic peptide family, which is involved in the control of extracellular fluid volume and electrolyte

homeostasis. Similarly, the protein NPPB's biological functions include natriuresis, diuresis, as well as the inhibition of renin and aldosterone secretion thus potentially playing a role in regulating blood pressure and cardiovascular function. Also, the adjacent gene *AGTRAP* encodes for the angiotensin II receptor associated protein, which is postulated to be involved in the regulation of blood pressure. Recent studies suggest that AGTRAP, or ATRAP, which is expressed in heart tissue plays a regulatory role of the cardiac SERCA2a uptake of calcium, and that it is associated with long-term inhibition of ventricular hypertrophy when hypomethylated [26, 27].

We used data from the BIOS QTL browser to assess whether differential methylation at the *AGTRAP-MTHFR-NPPA/B* gene cluster was functionally associated with variability in expression of any genes in *cis* of this gene cluster (±500kb)[14]. Our *in-silico* look up identified the expression of the *TNFRSF8* gene on chromosome 1 to be significantly associated with differential methylation of the probe cg06193043 probe at this gene cluster (P= 1.5E-04). *TNFRSF8* encodes a protein of the tumor necrosis factor-receptor superfamily (CD30) that mediates the signal transduction pathway leading to NF-kB activation and so is a positive regulator of apoptosis. Expression of CD30 has been previously associated with autoimmune function [28] and cancer therapy [29] but no over-obvious link can be drawn to 1-CM function.

DMPs in HIF3A are associated with aspects of metabolic syndrome

The CpG sites at loci on chromosomes 2 (*PLEKHH2*), 7 (*SH2B2*) and 19 (*HIF3A*) were shown to be significantly associated with [*c.677C>T* x tHcy] levels in both cohorts and thought to play roles in aspects of metabolic syndrome. HIF3A is thought to act as a transcriptional regulator in adaptive response to low oxygen tension and plays a role in glucose, lipid and amino acid metabolism as well as adipocyte differentiation [30, 31]. Interestingly, hyper-methylation at the *HIF3A* locus has been previously positively associated with variability in BMI [32], adipose tissue dysfunction [33], and risk of gestational diabetes mellitus[34]. There have also been speculations about the role of genetic variants in the *PLEKHH2* region as a potential risk factor for diabetic neuropathy [35]. To date, no human studies have been carried out to investigate site-specific DNA methylation patterns related to 1-CM that may influence risk of obesity and metabolic syndrome. There is compelling evidence in mouse models to suggest a role of 1-CM in mediating the risk of metabolic syndrome through altered DNA methylation and how the provision of methyl donor supplements can avert transgenerational effects of obesity [36]. There is however robust observational data both supporting or refuting an observational link between 1-CM function – through studies on tHcy, folate, or vitamin B_{12} as well as genetic studies, based on either the *MTHFR* polymorphism alone [37] or a cassette of genetic markers [38] - and aberrant metabolic function [39, 40], obesity, or type 2 diabetes [41, 42]. Results from the current study add further evidence to this literature and motivate future functional

work on these identified loci which may allow more robust conclusions to be drawn with regard to the role of 1-CM function in metabolic disease aetiology [34, 43, 44].

DMPs in genes associated with cellular structure/function and cancer

Finally, we also identified significant associations with CpG sites on chromosome 2 (located within/near genes *DYNC2LI1, NCAPH, NEURL3*) and 4 (*PROM1*) which are thought to play putative roles in cellular structure and function. NCAPH has been found to be highly expressed in colon cancerous (CC) tissue and harbours a number of mutations in CC patients, serving as a prime therapeutic target for CC [45]. Similarly, *PROM1* (*CD133*) is associated with recurrent glioma [46] and melanoma [47]. Due to their involvement in cellular differentiation and transport, we can hypothesise that alterations in the expression of these genes due to epigenetic modifications caused by 1-CM imbalance may lead to cell metabolism imbalances, potentially leading to a number of complications, including cellular atrophy, neurodegeneration and different types of cancers.

Strengths and limitation

Our study has specific strengths and limitations which warrant further discussion. Our study is the first to systematically investigate site-specific DNA methylation as a consequence of interaction between the *MTHFR c.677C>T* polymorphism and tHcy. Furthermore, the utility of two independent cohorts followed by a meta-analysis reduces potential biases, the chance of false positive findings whilst possibly increasing our power to detect loci of significant biological effect. Finally, utilising an established MWAS analytical approach incorporating and correcting for technical and biological covariates into our analysis provides reassurance that our results represent genuine findings.

A potential limitation of our study is the unbalanced constituency of gender among our cohorts. For historical reasons, the TwinsUK cohort is predominantly made up of female twins whereas the Rotterdam cohort included a balanced number of males and females. Thus, inherent male/female differences in either the exposure (tHcy) and/or outcome (site-specific DNA methylation levels) may distort the molecular consequences to the discordant tHcy levels in males and females. Such gender differences have been identified in tHcy levels previously. Total Hcy concentrations have been found to be significantly higher in men than in women in both the Framingham Offspring Cohort study and the Hordaland Homocysteine study which demonstrated ~11% higher tHcy in men than women[48, 49]. In our study, we also note a significant difference in tHcy levels between males and females in the Rotterdam cohort matching those reported in the HHS (~11%). To overcome this limitation, we included sex as a covariate in the Rotterdam MWAS and notably we identified some of the same significant loci in both the female-only meta-analysis and

the full cohort (male + females) meta-analysis. Another potential limitation is that DNA was extracted from whole blood and so the site-specific methylation values represent those from across various types of leukocytes. To minimise the influence of cell composition we adjusted our analysis for either measured or estimated leukocytes composition but there is still a possibility that observed differential methylation was confounded by either tHcy or ageing induced shifts in leukocyte distribution. Finally, although we used the latest technology for assessing DNA methylation patterns from across the human genome, the450 BeadChip array still only targets ~2% of all CpGs in the genome. Thus, although not strictly a shortcoming of our study, it is quite likely that we would have missed a number of relevant genomic regions.

CONCLUSIONS

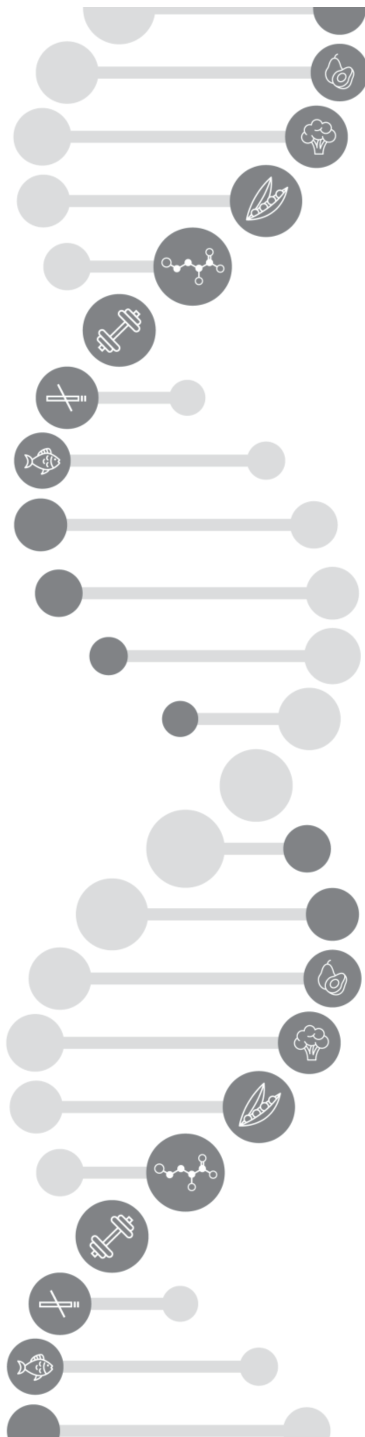
This study has allowed us to test and further examine an important paradigm that there are associations between DNA methylation patterns and [*c.677C>T* x tHcy] at specific sites in the epigenome, and to gain novel potential mechanistic insight into the epigenetic consequences of this gene by nutrient/metabolite interaction. Indeed, identifying [*c.677C>T* x tHcy]-sensitive DMPs, and their effect on nearby genes has provided a step forward in (1) identifying loci that can be used as candidate genes or regions in specific diseases, and (2) generating opportunities for better/well-informed study design, driving us towards more targeted preventative public health approach.

REFERENCES

1. Carmel R, Jacobsen DW, *Homocysteine in Health and Disease*. 2001, Cambridge, UK.
2. Mandaviya PR, Stolk L, Heil SG. Homocysteine and DNA methylation: a review of animal and human literature. Mol Genet Metab 2014;113:243-52.
3. Clarke R, Smith AD, Jobst KA, Refsum H, Sutton L, Ueland PM. Folate, vitamin B12, and serum total homocysteine levels in confirmed Alzheimer disease. Arch Neurol 1998;55:1449-55.
4. Seshadri S, Beiser A, Selhub J, Jacques PF, Rosenberg IH, D'Agostino RB, et al. Plasma homocysteine as a risk factor for dementia and Alzheimer's disease. N Engl J Med 2002;346:476-83.
5. Wald DS, Kasturiratne A, Simmonds M. Serum homocysteine and dementia: Meta-analysis of eight cohort studies including 8669 participants. Alzheimers Dement 2011;7:412-7.
6. Rampersaud GC, Kauwell GP, Hutson AD, Cerda JJ, Bailey LB. Genomic DNA methylation decreases in response to moderate folate depletion in elderly women. Am J Clin Nutr 2000;72:998-1003.
7. Ingrosso D, Cimmino A, Perna AF, Masella L, De Santo NG, De Bonis ML, et al. Folate treatment and unbalanced methylation and changes of allelic expression induced by hyperhomocysteinaemia in patients with uraemia. Lancet 2003;361:1693-9.
8. Fryer AA, Nafee TM, Ismail KM, Carroll WD, Emes RD, Farrell WE. LINE-1 DNA methylation is inversely correlated with cord plasma homocysteine in man: a preliminary study. Epigenetics 2009;4:394-8.
9. Dominguez-Salas P, Cox SE, Prentice AM, Hennig BJ, Moore SE. Maternal nutritional status, C(1) metabolism and offspring DNA methylation: a review of current evidence in human subjects. Proc Nutr Soc 2012;71:154-65.
10. Friso S, Choi SW, Girelli D, Mason JB, Dolnikowski GG, Bagley PJ, et al. A common mutation in the 5,10-methylenetetrahydrofolate reductase gene affects genomic DNA methylation through an interaction with folate status. Proc Natl Acad Sci U S A 2002;99:5606-11.
11. Verhoef P, Kok FJ, Kluijtmans LA, Blom HJ, Refsum H, Ueland PM, Kruyssen DA. The 677C-->T mutation in the methylenetetrahydrofolate reductase gene: associations with plasma total homocysteine levels and risk of coronary atherosclerotic disease. Atherosclerosis 1997;132:105-13.
12. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, Beck S. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. Bioinformatics 2013;29:189-96.
13. Wilhelm-Benartzi CS, Koestler DC, Karagas MR, Flanagan JM, Christensen BC, Kelsey KT, et al. Review of processing and analysis methods for DNA methylation array data. Br J Cancer 2013;109:1394-402.
14. Bonder MJ, Luijk R, Zhernakova DV, Moed M, Deelen P, Vermaat M, et al. Disease variants alter transcription factor levels and methylation of their binding sites. Nat Genet 2017;49:131-138.
15. Zhernakova DV, Deelen P, Vermaat M, van Iterson M, van Galen M, Arindrarto W, et al. Identification of context-dependent expression quantitative trait loci in whole blood. Nat Genet 2017;49:139-145.
16. Castro R, Rivera I, Ravasco P, Camilo ME, Jakobs C, Blom HJ, de Almeida IT. 5,10-methylenetetrahydrofolate reductase (MTHFR) 677C-->T and 1298A-->C mutations are associated with DNA hypomethylation. J Med Genet 2004;41:454-8.
17. Stover PJ. One-carbon metabolism-genome interactions in folate-associated pathologies. J Nutr 2009;139:2402-5.

18. Schalinske KL, Smazal AL. Homocysteine imbalance: a pathological metabolic marker. Adv Nutr 2012;3:755-62.
19. Axume J, Smith SS, Pogribny IP, Moriarty DJ, Caudill MA. The MTHFR 677TT genotype and folate intake interact to lower global leukocyte DNA methylation in young Mexican American women. Nutr Res 2007;27:1365-1317.
20. Stern LL, Mason JB, Selhub J, Choi SW. Genomic DNA hypomethylation, a characteristic of most cancers, is present in peripheral leukocytes of individuals who are homozygous for the C677T polymorphism in the methylenetetrahydrofolate reductase gene. Cancer Epidemiol Biomarkers Prev 2000;9:849-53.
21. Shelnutt KP, Kauwell GP, Gregory JF, 3rd, Maneval DR, Quinlivan EP, Theriaque DW, et al. Methylenetetrahydrofolate reductase 677C-->T polymorphism affects DNA methylation in response to controlled folate intake in young women. J Nutr Biochem 2004;15:554-60.
22. Amarasekera M, Martino D, Ashley S, Harb H, Kesper D, Strickland D, et al. Genome-wide DNA methylation profiling identifies a folate-sensitive region of differential methylation upstream of ZFP57-imprinting regulator in humans. FASEB J 2014;28:4068-76.
23. Kok DE, Dhonukshe-Rutten RA, Lute C, Heil SG, Uitterlinden AG, van der Velde N, et al. The effects of long-term daily folic acid and vitamin B12 supplementation on genome-wide DNA methylation in elderly subjects. Clin Epigenetics 2015;7:121.
24. Mandaviya PR, Aissi D, Dekkers KF, Joehanes R, Kasela S, Truong V, et al. Homocysteine levels associate with subtle changes in leukocyte DNA methylation: an epigenome-wide analysis. Epigenomics 2017;9:1403-1422.
25. Mandaviya PR, Joehanes R, Aissi D, Kuhnel B, Marioni RE, Truong V, et al. Genetically defined elevated homocysteine levels do not result in widespread changes of DNA methylation in leukocytes. PLoS One 2017;12:e0182472.
26. Mederle K, Gess B, Pluteanu F, Plackic J, Tiefenbach KJ, Grill A, et al. The angiotensin receptor-associated protein Atrap is a stimulator of the cardiac Ca2+-ATPase SERCA2a. Cardiovasc Res 2016;110:359-70.
27. Wang TJ, Lian GL, Lin X, Zhong HB, Xu CS, Wang HJ, Xie LD. Hypomethylation of Agtrap is associated with long-term inhibition of left ventricular hypertrophy in prehypertensive losartan-treated spontaneously hypertensive rats. Mol Med Rep 2017;15:839-846.
28. Shinoda K, Sun X, Oyamada A, Yamada H, Muta H, Podack ER, et al. CD30 ligand is a new therapeutic target for central nervous system autoimmunity. J Autoimmun 2015;57:14-23.
29. Muta H, Podack ER. CD30: from basic research to cancer therapy. Immunol Res 2013;57:151-8.
30. Hatanaka M, Shimba S, Sakaue M, Kondo Y, Kagechika H, Kokame K, et al. Hypoxia-inducible factor-3alpha functions as an accelerator of 3T3-L1 adipose differentiation. Biol Pharm Bull 2009;32:1166-72.
31. Sheng L, Liu Y, Jiang L, Chen Z, Zhou Y, Cho KW, Rui L. Hepatic SH2B1 and SH2B2 regulate liver lipid metabolism and VLDL secretion in mice. PLoS One 2013;8:e83269.
32. Dick KJ, Nelson CP, Tsaprouni L, Sandling JK, Aissi D, Wahl S, et al. DNA methylation and body-mass index: a genome-wide analysis. Lancet 2014;383:1990-8.
33. Pfeiffer S, Kruger J, Maierhofer A, Bottcher Y, Kloting N, El Hajj N, et al. Hypoxia-inducible factor 3A gene expression and methylation in adipose tissue is related to adipose tissue dysfunction. Sci Rep 2016;6:27969.
34. Haertle L, El Hajj N, Dittrich M, Muller T, Nanda I, Lehnen H, Haaf T. Epigenetic signatures of gestational diabetes mellitus on cord blood methylation. Clin Epigenetics 2017;9:28.

35. Greene CN, Keong LM, Cordovado SK, Mueller PW. Sequence variants in the PLEKHH2 region are associated with diabetic nephropathy in the GoKinD study population. Hum Genet 2008;124:255-62.
36. Waterland RA, Dolinoy DC, Lin JR, Smith CA, Shi X, Tahiliani KG. Maternal methyl supplements increase offspring DNA methylation at Axin Fused. Genesis 2006;44:401-6.
37. Zhong JH, Rodriguez AC, Yang NN, Li LQ. Methylenetetrahydrofolate reductase gene polymorphism and risk of type 2 diabetes mellitus. PLoS One 2013;8:e74521.
38. Kumar J, Ingelsson E, Lind L, Fall T. No Evidence of a Causal Relationship between Plasma Homocysteine and Type 2 Diabetes: A Mendelian Randomization Study. Front Cardiovasc Med 2015;2:11.
39. Buysschaert M, Dramais AS, Wallemacq PE, Hermans MP. Hyperhomocysteinemia in type 2 diabetes: relationship to macroangiopathy, nephropathy, and insulin resistance. Diabetes Care 2000;23:1816-22.
40. Patterson S, Flatt PR, Brennan L, Newsholme P, McClenaghan NH. Detrimental actions of metabolic syndrome risk factor, homocysteine, on pancreatic beta-cell glucose metabolism and insulin secretion. J Endocrinol 2006;189:301-10.
41. Hoogeveen EK, Kostense PJ, Jakobs C, Dekker JM, Nijpels G, Heine RJ, et al. Hyperhomocysteinemia increases risk of death, especially in type 2 diabetes : 5-year follow-up of the Hoorn Study. Circulation 2000;101:1506-11.
42. Meigs JB, Jacques PF, Selhub J, Singer DE, Nathan DM, Rifai N, et al. Fasting plasma homocysteine levels in the insulin resistance syndrome: the Framingham offspring study. Diabetes Care 2001;24:1403-10.
43. Main AM, Gillberg L, Jacobsen AL, Nilsson E, Gjesing AP, Hansen T, et al. DNA methylation and gene expression of HIF3A: cross-tissue validation and associations with BMI and insulin resistance. Clin Epigenetics 2016;8:89.
44. Mok A, Solomon O, Nayak RR, Coit P, Quach HL, Nititham J, et al. Genome-wide profiling identifies associations between lupus nephritis and differential methylation of genes regulating tissue hypoxia and type 1 interferon responses. Lupus Sci Med 2016;3:e000183.
45. Yin L, Jiang LP, Shen QS, Xiong QX, Zhuo X, Zhang LL, et al. NCAPH plays important roles in human colon cancer. Cell Death Dis 2017;8:e2680.
46. Sun B, Wan Z, Shen J, Ni L, Chen J, Cui M, et al. DNA hypomethylation of CD133 promoter is associated with recurrent glioma. Oncol Rep 2016;36:1062-8.
47. Madjd Z, Erfani E, Gheytanchi E, Moradi-Lakeh M, Shariftabrizi A, Asadi-Lari M. Expression of CD133 cancer stem cell marker in melanoma: a systematic review and meta-analysis. Int J Biol Markers 2016;31:e118-25.
48. Nygard O, Refsum H, Ueland PM, Vollset SE. Major lifestyle determinants of plasma total homocysteine distribution: the Hordaland Homocysteine Study. Am J Clin Nutr 1998;67:263-70.
49. Jacques PF, Bostom AG, Wilson PW, Rich S, Rosenberg IH, Selhub J. Determinants of plasma total homocysteine concentration in the Framingham Offspring cohort. Am J Clin Nutr 2001;73:613-21.



PART C

Nutrition, Lifestyle
And DNA methylation

CHAPTER 7

Association of dietary folate and vitamin B12 intake with genome-wide DNA methylation; a large scale epigenome-wide association analysis in 5,841 individuals

Pooja R. Mandaviya*, Roby Joehanes*, Jennifer Brody*, Juan E. Castillo-Fernandez*, Koen K. F. Dekkers*, Anh N. Do*, Mariaelisa Graff*, Ismo K. Hänninen*, Toshiko Tanaka*, Ester A.L. de Jonge, Jessica C. Kiefte-de Jong, Devin M. Absher, Stella Aslibekyan, Yolanda B. de Rijke, Myriam Fornage, Dena G. Hernandez, Mikko A. Hurme, M. Arfan Ikram, Paul F. Jacques, Anne E. Justice, Douglas P. Kiel, Rozenn N Lemaitre, Michael M. Mendelson, Vera Mikkilä, Ann Z. Moore, Tess Pallister, Olli T. Raitakari, Casper G. Schalkwijk, Jin Sha, Eline P. E. Slagboom, Caren E. Smith, Coen D.A. Stehouwer, Pei-Chien Tsai, André G. Uitterlinden, Carla J.H. van der Kallen, Diana D. van Heemst, BIOS Consortium, CHARGE Consortium Epigenetics group, CHARGE Consortium Nutrition group, Donna K. Arnett[#] [28], Stefania Bandinelli[#], Jordana T. Bell[#], Bastiaan B. T. Heijmans[#], Terho Lehtimäki[#], Daniel Levy[#], Kari E. North[#] [9], Nona Sotoodehnia[#] [5], Marleen M.J. van Greevenbroek[#], Joyce B.J. van Meurs[#], Sandra G. Heil[#]

These authors contributed equally as first authors

***These authors contributed equally as senior authors*

Submitted for publication

ABSTRACT

Background: Folate and vitamin B12 are essential micronutrients involved in the donation of methyl groups in cellular metabolism. Folate and vitamin B12 intake have been shown to increase DNA methylation in mostly animal studies. However, associations between intake of these nutrients and genome-wide DNA methylation levels have not been studied comprehensively in humans.

Objective: To assess whether folate or vitamin B12 intake are asssociated with genome-wide changes in methylation in leukocytes.

Design: A large-scale epigenome-wide association study of folate and vitamin B12 intake was performed on DNA from 5,841 participants from 10 cohorts using Illumina 450k arrays. Folate and vitamin B12 intakes were calculated from food frequency questionaires. Continuous and categorical (low vs. high intake) linear regression mixed models were applied per cohort, controlling for confounders. Meta-analysis was performed to identify significant differentially methylated positions (DMPs) and regions (DMRs). Pathway analysis was done on the DMR annotated genes using the WEB-based GEne SeT AnaLysis Toolkit.

Results: Folate intake was associated with 6 novel DMPs annotated to *FAM64A*, *WRAP73*, *FRMD8*, *CUX1* and *LCN8* genes, which have a role in cellular processes including centrosome localization, cell proliferation and tumorigenesis. Regional analysis showed 74 folate associated DMRs annotated to 117 genes. The most significant DMR was a 400 bp spanning region annotated to *LGALS3BP* that is implicated in immune response and tumorigenesis. Signaling pathways regulating pluripotency of stem cells are overrepresented among the folate associated DMR genes. Vitamin B12 intake was associated with 29 DMRs annotated to 48 genes, of which the most significant was a 1,100 bp spanning region annotated to *CABYR,* which is involved in lung and hepatocellular cancers. Vitamin B12 intake was not associated with DMPs.

Conclusions: We identified several novel differentially methylated loci that could be mechanistic indicators of low folate and vitamin B12 intake.

INTRODUCTION

Folate and vitamin B12 are essential micronutrients of the one-carbon pathways that are involved in the donation of methyl groups to the DNA, RNA and proteins [1]. Folate is a methyl-donor itself, where its active form, 5-methyltetrahydrofolate, donates its methyl group to the re-methylation of homocysteine to methionine. Vitamin B12 is a cofactor in this reaction.

Folate and vitamin B12 status has been connected to diverse diseases. Low levels of folate and vitamin B12 during pregnancy are independently associated with the risk of neural tube defects in the child [2-4]. IN addition, low levels of folate and/or vitamin B12 are associated with risk for a wide range of diseases, such as cardiovascular diseases and osteoporosis [5-8]. Severe vitamin B12 and folate deficiencies result in megaloblastic anemia, which for vitamin B12 deficiency is associated with severe neurological abnormalities [9]. Diagnosis of vitamin B12 deficiency is hampered by the lack of diagnostic accuracy of available biomarkers. DNA methylation is a possible mechanism underlying the previously identified relationships between folate or vitamin B12 deficiency and disease risk, and specific alterations in methylation patterns could serve as future biomarkers for these nutrient related disease risks. Therefore, it is important to assess the association of these nutrients with DNA methylation.

The relationship between folate and/or vitamin B12 intake and DNA methylation has been investigated mostly in studies examining individuals with a particular disease, which might confound the observed associations [10-12]. A limited number of human studies were conducted in disease-free individuals [13-17]. Importantly, in all previous studies, global DNA methylation levels were assessed either as total 5-methyl cytosine content, or LINE-1 or Alu repeat methylation as a proxy for global methylation status in blood leukocytes. The results from these studies were inconsistent [13-17]. The relation between maternal folate intake and fetal DNA methylation was investigated in a few studies showing that maternal intake was associated with changes in DNA methylation [18-20]. The previous studies related to folate and vitamin B12 and DNA methylation have been limited with respect to outcome (e.g. global methylation levels only), sample size and participants (e.g. maternal-fetal exposures), and to date there are no large-scale epigenome-wide association studies (EWASs) in adults examining the relationship between B-vitamin intake and DNA methylation.

In this study, we undertook a large-scale EWAS of folate and vitamin B12 intake, analysing the association with methylation at up to 485,512 CpGs assessed in whole blood measured in up to 5,841 individuals across ten cohorts from Europe and North America. Since folate and vitamin B12 are important in the transfer of methyl groups to DNA, we hypothesize that low intake of folate or vitamin B12 is associated with genome-wide DNA hypomethylation in the normal population.

MATERIALS AND METHODS

Study populations

Data from 10 cohorts with a total of 5,520 individuals of European ancestry and 321 African American individuals were included in this meta-analysis. Written informed consent was given by all participants for genetic research. Descriptions of each cohort are provided in Text S1. For all participating studies, individuals with prevalent cancer were excluded from analyses due to potential differences in dietary patterns in response to their disease [21, 22] and different methylation patterns [23, 24].

DNA methylation assessment

Genomic methylation profiling was performed on whole blood in 10 cohorts using the Infinium Illumina HumanMethylation 450k BeadChip arrays (Illumina Inc., San Diego, USA) according to the manufacturer's protocol. The Genetics of Lipid Lowering Drugs and Diet Network (GOLDN) study used $CD4^{+}$ T cells from buffy coats. The Illumina array measures methylation status of 485,512 CpG sites in the gene and non-gene regions of CpG islands, shores and shelves of the human genome [25]. Poor quality samples and probes were excluded based on cohort-specific criteria [Table S1]. Quantile [26], DASEN [27], Subset-quantile Within Array Normalization (SWAN) [28], Beta MIxture Quantile dilation (BMIQ) [29] or Functional normalization [30] was used to correct the raw beta values that represents the methylation percentage per CpG for every sample. The normalized beta values were used for the association analysis in each cohort.

Data collection and Dietary assessment

Dietary intake data were derived in each cohort from structured self-administered food frequency questionnaires (FFQs) varied from 78 to 560 food items [Table 1]. Exposure variables for the nutrients folate and vitamin B12 were calculated in micrograms per day from dietary intake data using national food composition tables [Table 1]. This included foods with folic acid fortification if this was the case, but did not include B-vitamin supplement intake. Participants with missing dietary data, or who reported very low (<500) or very high (>5000) total energy intake were excluded.

In order to reduce the magnitude of the systematic measurement error of the FFQ, each nutrient was adjusted for total energy intake, using the residual method [31]. Next, unstandardized residuals from this regression of each nutrient and total energy intake were used for association analyses. B-vitamin supplement intake was available in all cohorts for use as a covariate. Since these supplement data were recorded in different forms (B-vitamins, multivitamins or folic acid supplements) and different units (frequency per day or per week) in each cohort, we harmonized these data across all cohorts by grouping individuals as supplement users and non-users.

STATISTICAL ANALYSES

Differentially Methylated Positions (DMPs): Cohort-specific association analyses

Each participating cohort used linear mixed models to investigate the associations between each nutrient and genome-wide DNA methylation. Models were adjusted for age, sex, body mass index (BMI), differential white blood cell (WBC) counts, smoking status, physical activity, B-vitamin supplement intake, and alcohol (grams per day or drinks per week) and coffee (grams per day or servings per day) consumption as fixed effects, except for those studies for whom some covariates were not present [Table 2]. Technical covariates such as array number and position on array were also adjusted for, and were treated as random effects. Differential WBC counts were either represented as percentage of measured cell counts, or imputed using the Houseman method [32]. We first used a continuous model (Model 1), where the unstandardized residuals of each nutrient were used as continuous variables to estimate their association with genome-wide DNA methylation. To determine whether large differences in nutrient intake effect DNA methylation, we also performed a categorical analysis where the individuals of each residual nutrient were divided into tertiles, and the first and third tertiles were used to define low and high nutrient intakes, respectively (Model 2). Food intake measured by FFQ is prone to measurement error, but is ideal for ranking individuals based on their intake [33]. Thus, a model based on categorical ranks rather than absolute intake is the recommended analytical approach. In addition, we performed a sensitivity analysis of both folate and vitamin B12 intake in non-users of B-vitamin supplements only (Model 3).

Differentially Methylated Positions (DMPs): Meta-analyses

Figure 1 shows the stepwise study design that was followed. Using the software GWAMA [34], fixed effect meta-analyses as weighted by inverse variance were performed from the summary statistics of each participating cohort on the continuous (Model 1), categorical (Model 2) and supplement non-users sensitivity models (Model 2a) of folate and vitamin B12 intake EWASs. An additional sensitivity meta-analysis was also performed with limitation to studies that had FFQ data at the same time point as DNA methylation measurements (Model 2b). Heterogeneity (I^2) was considered to account for differences between cohorts using fixed effect meta-analysis. The probes with SNPs at single base extension and probes with improper binding [35] were excluded to avoid spurious signals and co-hybridization with alternate homologous sequences. All participating cohorts had different probe exclusions for quality control and therefore at this stage, we removed probes if they were not present in at least 5 cohorts. The significance was defined by the Benjamini-Hochberg (BH) method [36] of the false discovery rate (FDR) <0.05. The gene annotations for the DMPs we identified were performed using the Genomic Regions Enrichment of Annotations Tool (GREAT)

[37] with University of California, Santa Cruz (UCSC) [38] where they assigned a regulatory domain consisting of a basal domain to each gene, that extends up to 5 kb upstream and 1 kb downstream from its transcription start site. The DMP is annotated with a gene if it overlaps with its basal domain, In addition, DMP is annotated to a gene if an extension is reached up to the basal regulatory domain of the nearest upstream and downstream genes within 1 Mb. The genomic inflation factor (λ) [39] were computed to estimate the rate of false positives due to population structure. Although such measure has been successfully applied in genome-wide association studies (GWAS), its application to EWAS may not confer the same benefit due to the inherent nature of the correlation structure in CpG sites of interdependent pathways and environment exposure [40].

Differentially Methylated Regions (DMRs)

We identified DMRs for each nutrient using the software comb-p [41], through analysis of the nominal p-values of DMPs generated from the meta-analysis (Figure 1, Model 3). Nominal p-values of DMPs were adjusted according to their weighted correlation with adjacent p-values using the Stouffer–Liptak–Kechris method, in a sliding window of 500 base pairs (bp) with varying lags of 50 bp [41]. Regions were identified by the peak finding algorithm on the adjusted p-values that qualified according to the BH FDR threshold <0.05. The identified regions were then given new p-values and corrected for multiple testing using the Sidak correction [41]. The significance for DMRs was defined as Sidak <0.05 for each nutrient. The gene annotations for the identified DMRs were performed using the GREAT tool [37] with the same extensions as applied for the DMP gene annotations.

Pathway analysis was performed for genes annotated to the DMRs related to each nutrient, using the WEB-based GEne SeT AnaLysis Toolkit [42]. Overrepresentation enrichment analysis was performed for the GREAT annotated genes of the folate and vitamin B12 associated DMRs, respectively, and compared against the reference genome, in the KEGG pathways. The multiple testing was set to FDR <0.05 and the minimum number of genes in a functional gene set category was set to two.

Previously published DMPs using Illumina 450k arrays

We also investigated whether previously published DMPs related to maternal folate [19, 20] and DMPs related to folate and vitamin B12 intervention [43] were associated with folate and vitamin B12 intake in our population-based meta-analysis in adults. Gonseth et al. observed four DMPs (cg22664307, cg21039708, cg15219145 and cg13499966) in neonatal blood which were associated with maternal folate intake (n=167) [19]. In a second study examining cord blood methylation with maternal plasma folate (n=1988), 443 DMPs in cord blood were identified [20]. In a 2-year folic

acid and vitamin B12 intervention study, one DMP (cg19380919) had the greatest change in methylation in the treatment group compared to the controls with marginal significance [43]. We compared the nominal p-values and beta coefficients to examine the direction of effects. In addition, we also checked for the enrichment of these previously found DMPs in our meta-analysis by comparing these CpGs to the same number of randomly selected CpGs from the array using 100 permutations. From both the CpG sets, the number of CpGs that had nominal p-value <0.05 were determined and compared. Using Fisher's exact test, the significant enrichment was then determined.

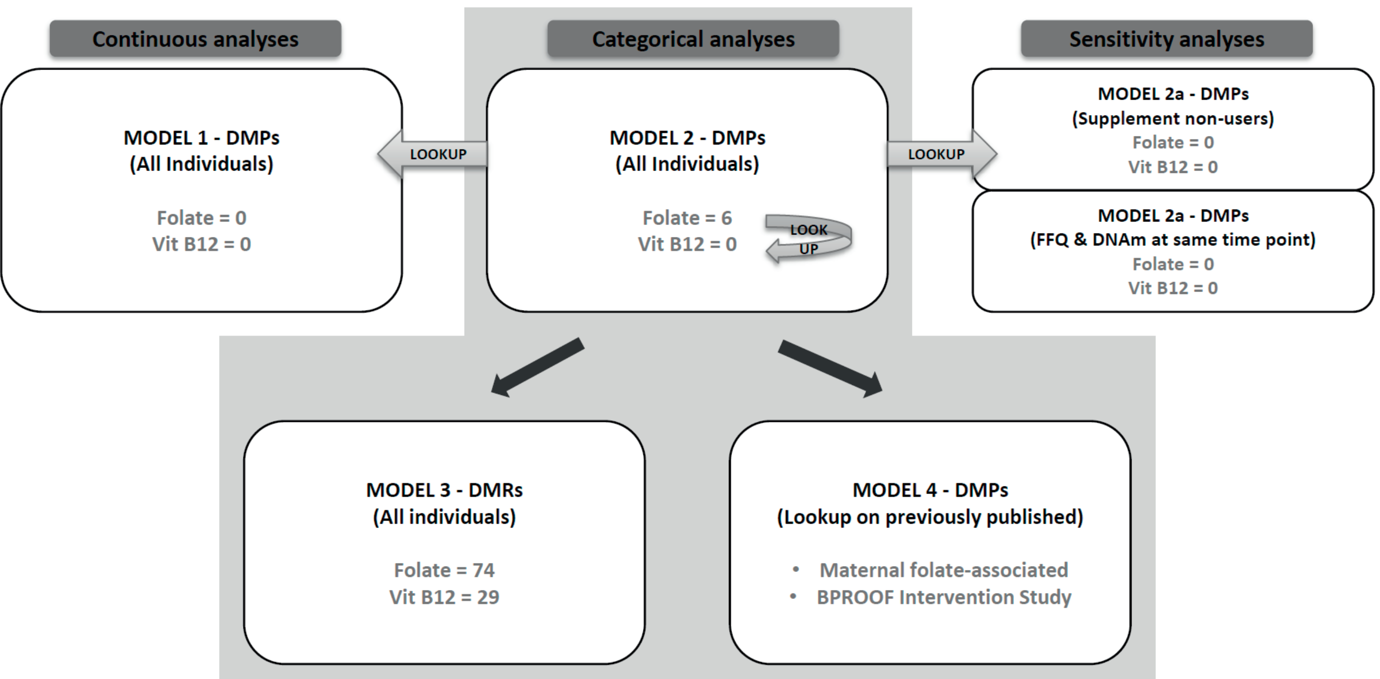


Figure 1. Analysis flow scheme. (1) Meta-analysis of continuous model EWAS of folate and vitamin B12 intake as well as (2) meta-analysis of categorical model EWAS were performed. Significant (FDR < 0.05) hits of the folate intake categorical meta-analysis were looked up in the folate intake continuous meta-analysis and vitamin B12 intake categorical meta-analysis. Meta-analyses of categorical EWAS were followed by (2a) sensitivity meta-analyses of B-vitamin supplement non-users which included lookup of the folate intake associated CpGs of the main model, (2b) sensitivity meta-analysis of studies with FFQ collection at same time point as DNA methylation measurements, (3) DMR analyses to find regions of association, and lastly, (4) comparison to previously identified DMPs associated with maternal folate intake.

RESULTS

Cohort Characteristics

Among the European cohorts, the median intakes of dietary folate ranged from 193 µg/d in the Leiden Longevity Study to 415 µg/d in the Rotterdam Study. Among the American cohorts, the median intakes of dietary folate ranged from 234 µg/d in the Genetics of Lipid Lowering Drugs and Diet Network (GOLDN) study to 405 µg/d in the Framingham Heart Study (FHS). Among the European cohorts, the median intake of dietary vitamin B12 ranged from 4.0 µg/d in the Leiden Longevity Study (LLS) to 7.4 µg/d in the Young Finns Study (YFS). For the American cohorts, the median intake of dietary vitamin B12 ranged from 4.5 µg/d in the GOLDN study to 6.3 µg/d in the ARIC Study. Characteristics of each participating study are given in Table 1.

Study participants were adults (44-100% women) with the mean age ranging from 40-77 years. There were 7-25% current smokers and 3-58% of B-vitamin, multivitamin or folic acid supplement users. Since the use of supplements may be a confounder to the analysis, we also performed a sensitivity analysis in non-users of B-vitamin, multivitamin or folic acid supplements. Further characteristics such as BMI, WBC counts, physical activity, alcohol and coffee intake of each cohort are provided in Table 2.

MODELS

Model 1: Continuous (Differentially methylated positions)

Meta-analysis of the 5,815 individuals from the EWAS analysis of the folate intake continuous model showed no significant DMP associations (FDR <0.05). Meta-analysis of the 5,302 individuals of the EWAS analysis of the vitamin B12 intake continuous model also showed no significant DMP associations (FDR <0.05).

Model 2: Categorical (Differentially methylated positions)

Meta-analyses of 3,894 individuals evaluated categorically from the EWASs of folate intake showed 6 significant DMP associations (FDR <0.05) [Table 3, Figure 2 (a-g) and S2]. These 6 DMPs showed consistent association with nominal significance (p <0.05) in the continuous model [Table 3]. The most significant DMP was at cg23465990 (P=3.87E-08, FDR=0.018) on chromosome 17, annotated to the nearest gene *FAM64A* (605 bp upstream), and showed 0.12% decrease in methylation per micrograms per day increase in residual folate intake. Other significant DMPs at cg11832534, cg03249011, cg14398883, cg00826902 and cg14145338 were annotated to the nearest genes *WRAP73* (2,648 bp downstream), *FRMD8* (41931 bp downstream), *CUX1* (62,673 bp upstream), *WRAP73* (2,692 bp downstream) and *LCN8* (3,667 bp downstream), respectively of chromosomes 1, 11, 7 and 9. All identified DMPs were negatively

associations between folate intake and methylation levels with 0.12-0.79% decrease in methylation per micrograms per day increase in residual folate intake.

EWAS results of the vitamin B12 intake categorical model did not show significant DMPs in the meta-analysis of 3,566 individuals [Figure S2]. Of the 6 significant DMPs from the folate intake analysis, two DMPs (cg23465990 and cg14398883) showed borderline nominal significance in the same direction as folate intake in the categorical model of vitamin B12 intake (p = 0.02 and 0.04, respectively) [Table S3].

Model 2a: Sensitivity model in supplement non-users

A sensitivity analysis was further performed to reduce heterogeneity in the models caused by the use of B-vitamin supplements. This may potentially help with finding new true positive DMPs. In sensitivity analyses performed in supplement non-users, there were no significant DMPs in the meta-analysis of 2,183 individuals for the categorical model of folate intake [Figure S3]. The 6 significant DMPs identified in the full categorical model were nominally associated (p <0.05) in the same direction with folate intake in the non-users [Table 3]. The effect sizes were similar to the full categorical model with 0.12-0.74% decrease in methylation per micrograms per day increase in residual folate intake. Furthermore, no significant DMPs were identified either in the individual cohort results nor the meta-analysis of 1,855 individuals for the categorical model of vitamin B12 intake.

Model 2b: Sensitivity model time of measurements

To determine whether there is a confounding effect by studies with a time-lag between DNA methylation and folate or vitamin B12 intake assessment, a second sensitivity meta-analysis was performed including only studies where DNA methylation measurements and FFQ collection were assessed at the same time. The FHS, YFS and a subset of the RS cohort had FFQ collection 4-7 years apart of DNA collection, and therefore, these three cohorts were left out in model 2b. Meta-analysis of the 2,183 individuals for the categorical model of folate intake did not result in significant DMPs [Figure S4]. The 6 significant DMPs associated with folate intake in the full categorical model had consistent associations with nominal significance [Table 3]. Furthermore, sensitivity meta-analysis of the 1,865 individuals for the categorical model of the vitamin B12 intake also did not identify significant DMPs.

Model 3: Categorical (Differentially methylated regions)

Folate intake

We additionally performed DMR analysis using p-values of the folate and vitamin B12 intake meta-analyses of the categorical model (Model 2). By investigation of significant regions (Sidak P<0.05) using the software comb-p, we observed 74 significant DMRs

associated with folate intake [Table 5, Figure 4a]. Most (73/74) DMRs were negatively associated with methylation. The most significant DMR associated with folate intake was the chr17:76,975,944-76,976,358 (P =1.47E-14) containing a 414 bp region of 8 DMPs, which annotates to the *LGALS3BP* gene. Overrepresentation enrichment analysis for the 117 genes of the folate associated 74 DMRs showed one significant pathway: signaling pathways regulating pluripotency of stem cells (P =1.61E-04).

Vitamin B12 intake

Regional analysis using comb-p found 29 significant DMRs with Sidak P<0.05 associated with vitamin B12 intake [Table 6, Figure 4b], of which 15 showed negative direction of effects. The most significant DMR associated with vitamin B12 intake was the chr18:21,718,458-21,719,569 (P =1.09E-13) containing a 1,111 bp region of 18 DMPs, and annotates to the promoter region of the gene *CABYR*. The vitamin B12 associated DMRs did not overlap with the folate intake associated DMRs. Enrichment analysis for the 48 genes of the vitamin B12 associated 29 DMRs did not reveal significant pathways.

PREVIOUSLY PUBLISHED DMPS USING ILLUMINA 450K ARRAYS

Maternal folate exposure and cord blood DMPs

We analysed previously identified DMPs that have previously been reported to be associated in neonates with maternal folate intake. In the current folate intake meta-analysis (Categorical Model 2), one (cg15219145) of the four previously identified DMPs from the Gonseth et al study [19] showed a nominal significance (P <0.05) with similar direction. None of these four CpGs were significant with vitamin B12 intake [Table S4]. Next, 28 and 27 of the 443 previously identified DMPs in newborns from the Joubert et al study [20] were associated with folate and vitamin B12 intake respectively in the same direction with nominal significance (P <0.05) in our study [Table S5]. However, no enrichment of significant p-values was found in these 443 DMPs in either folate (p = 0.60) or vitamin B12 (p = 0.22) intake models in our study.

Folic acid and Vitamin B12 Intervention study

The previously identified DMP (cg19380919) of a 2-year folic acid and vitamin B12 intervention study [43] was not associated with either folate intake (p = 0.78) or vitamin B12 intake (p = 0.43) in the current study [Table S6].

Table 1. Dietary assessment in participating cohorts.

Cohorts	COUNTRY	N (5,841)	Fortification	No. of Food items	Year of dietary data collection	Reference table	Tine gap (Years)
RS	NL	900	No	389	2006-2012	The Dutch Food Composition Tables (NEVO table, 2006)	0 (46% samples had ~7y gap)
LLS	NL	430	No	183	2006-2007	FQ13V20061221 (Wageningen 2006), calculations based on Dutch Food Composition Database (NEVO table, 2006)	0
CODAM	NL	154	No	178	2006-2009	NEVO table, 2001	0
InCHIANTI	IT	484	No	236	1998-2000	Italian Tables of Food Composition, 1998	0
YFS	FI	155	No	128	2007	Finnish Tables of Food Composition, 2007	4
TwinsUK	UK	568	Yes	131	1994-2001, 2007	McCance and Widdowson's The Composition of Foods Edition 6, 2002	0
FHS	USA	1,657	Yes	126	2005-2008	Harvard FFQ Nutrient Database, 2009	5-6
GOLDN	USA	983	Yes	124	2002-2004	The US Department of Agriculture's 1994–1996 Continuing Survey of Food Intakes by Individuals	0
ARIC	USA	321	No	78	1993-1995	USDA. Composition of foods: agriculture handbook no. 8 series. Washington, DC: US Department of Agriculture, 1975–1989	0
CHS	USA	188	No	131	1989-1990	Harvard FFQ Nutrient Database, 1996	0

N: Number of samples after exclusion of individuals with prevalent cancer and very low (<500) or very high (>5000) total energy intake
Fortification: Cohorts that included foods with folic acid fortification

Table 2. Demographic and lifestyle characteristics of participating cohorts.

Cohorts	Age [Mean (SD)]	Women [%]	BMI [Mean (SD)]	WBC Granulocytes [Mean (SD)]	WBC Lymphocytes [Mean (SD)]	WBC Monocytes [Mean (SD)]	Current smokers [%]	Alcohol intake [Mean (SD)]	Coffee intake [Mean (SD)]	Physical activity [Mean (SD)] Or Active [%]	Supp users [%]	Folate (µg/day) [Median (Range)]	Vit B12 (µg/day) [Median (Range)]
RS	64.5 (9.0)	57.4	27.6 (4.1)	0.4 (0.1)	0.1 (0.1), 0.2 (0.1) 0.1 (0.1), 0.1 (0.0)	0.1 (0.0)	7.4	12.3 (15.7)[a]	385.8 (240.0)[a]	63.4 (64.2)[a]	32.1	415.2 (53.9-1173.9)	4.6 (0.6-14.1)
LLS	58.6 (6.3)	51.3	25.3 (3.8)	4.3 (1.3)	2.0 (0.6)	0.4 (0.1)	12.3	16.7 (15.1)	NA	NA	26.3	192.7 (68.0-415.4)	4.0 (0.9-11.0)
CODAM	65.5 (6.9)	44.2	28.6 (4.2)	0.3 (0.1)	0.1 (0.1), 0.3 (0.1) 0.2 (0.1), 0.1 (0.0)	0.1 (0.0)	15.6	12.5 (14.2)[a]	491.9 (276.8)[a]	7150.4 (4955.3)[b]	3.2	200.3 (96.2-364.7)	4.2 (0.8-10.8)
InCHIANTI	62.6 (15.7)	54.1	27.1 (3.9)	0.031 (0.02)[a] 0.59 (0.08)[b]	0.32 (0.08)	0.05 (0.01)	18.6	16.5 (21.7)	2.4 (1.4)[b]	51	NA	273.9 (88.9-645.9)	NA
YFS	40.2 (3.3)	61.9	25.6 (4.4)	0.5 (0.1)	0.1 (0.0), 0.2 (0.1) 0.1 (0.1), 0.1 (0.0)	0.1 (0.0)	24.5	6.9 (7.3)	409.9 (265.8)[a]	77.4	39.4	332.8 (159.1-732.0)	7.4 (3.0-16.9)
TwinsUK	58.9 (9.9)	100.0	26.4 (4.8)	0.5 (0.1)	0.1 (0.0), 0.2 (0.1) 0.1 (0.0), 0.1 (0.0)	0.1 (0.0)	10.9	7.4 (11.2)	NA	NA	53.3	376.7 (122.6-976.3)	6.0 (0.5-16.0)
FHS	64.8 (8.6)	57.1	28.2 (5.4)	0.5 (0.1)	0.1 (0.1), 0.2 (0.1) 0.0 (0.0), 0.0 (0.0)	0.1 (0.0)	10.3	10.1 (15.4)	157.2 (123.8)[c]	88.3	57.7	405.9 (99.2-1141.0)	5.5 (0.0-18.2)
GOLDN	49.0 (16.0)	53.0	28.3 (5.7)	NAp	NAp	NAp	7.0	6.0 (20.0)	277.5 (427.1)	2.97 (2.47)[c]	58.0	233.6 (0.0-1143.6)	4.5 (0.0-31.2)
ARIC	59.5 (5.8)	67.6	30.5 (6.4)	1.7 (2.2)[a] 54.1 (10.2)[b]	NA	5.3 (3.4)	23.4	2.7 (10.0)	8.9 (12.1)	66.8 (52.0)[d]*	24.3	248.1 (36.0-811.2)	6.3 (0.7-26.3)
CHS	77.3 (4.9)	66.5	27.5 (4.8)	0.5 (0.15)	0.1 (0.1), 0.2 (0.1) 0.1 (0.1), 0.1 (0.0)	0.1 (0.04)	7.98	2.34 (6.6)	0.73 (1.1)[b]	763.9 (1188.1)	33.0	360.0 (93.7-886.7)	5.5 (0.8-28.7)

Measured cell counts in 10^9/L; Granulocytes: [a]Eosinophils and [b]Neutrophils, Houseman imputed percentage cell counts: CD8T, CD4T, NK, Bcell

Alcohol intake: [a]grams/day, Coffee intake: [a]grams/day or [b]servings/day; [c]FHS used caffeine intake in grams/day

Physical activity: [a]Total MET hours/week, [b]Total score of all activities, [c]Hours per week moderate activity, [d]Total MET minutes/week, [e]Kilocalories

NA: Not available, NAp: Not applicable, *Not included in analyses (only available to N=174 individual), Supp (Supplement)

Table 3. Differentially methylated positions significantly associated with folate intake at the epigenome-wide level in the meta-analysis.

CpG [Chr]	Nearest Gene	Categorical Model: All Individuals (Model 2; N=3,894)					Categorical Model: Sensitivity Analysis of Supplement Non-users LOOKUP (Model 2a; N=2,183)			Categorical Model: Sensitivity Analysis of Same time point FFQ & DNAm LOOKUP (Model 2b; N=2,193)			Continuous Model: All Individuals LOOKUP (Model 1; N=5,815)		
		Effect	SE	P	FDR	I^2	Effect	SE	P	Effect	SE	P	Effect	SE	P
cg23465990 [17]	*FAM64A* (-605)	-0.001	2.2E-04	3.9E-08	0.02	0.02	-0.001	2.6E-04	5.3E-06	-0.001	2.9E-04	9.2E-04	-4.0E-06	1.0E-06	2.4E-05
cg11832534 [1]	*WRAP73* (+2648), *TPRG1L* (+22467)	-0.004	7.7E-04	9.0E-08	0.02	0	-0.003	9.6E-04	7.9E-04	-0.004	9.7E-04	2.2E-05	-1.5E-05	3.0E-06	8.9E-06
cg03249011* [11]	*SCYL1* (-96576), *FRMD8* (+41931)	-0.008	1.4E-03	1.8E-07	0.03	0	-0.007	1.8E-03	4.2E-05	-0.008	1.9E-03	3.0E-05	-1.6E-05	6.0E-06	7.5E-03
cg14398883 [7]	*MYL10* (-125633), *CUX1* (-62673)	-0.008	1.6E-03	5.1E-07	0.05	0	-0.005	1.4E-03	5.2E-04	-0.007	2.2E-03	2.9E-03	-2.8E-05	6.0E-06	1.0E-06
cg00826902 [1]	*WRAP73* (+2692), *TPRG1L* (+22423)	-0.005	9.8E-04	5.9E-07	0.05	0.49	-0.005	1.2E-03	1.5E-05	-0.003	1.2E-03	3.4E-02	-1.9E-05	4.0E-06	3.2E-06
cg14145338 [9]	*LCN6* (-6084), *LCN8* (+3667)	-0.006	1.2E-03	6.2E-07	0.05	0.05	-0.006	1.4E-03	2.1E-05	-0.004	1.5E-03	6.2E-03	-2.2E-05	5.0E-06	2.0E-05

Effect: Beta coefficients based on unstandardized residuals of folate intake, adjusted for total energy intake

I^2: Heterogeneity I^2 parameter

Bp: Base pair location based on Illumina annotation

*Location=Enhancer, based on Illumina annotation, derived from UCSC

FDR threshold = 0.05

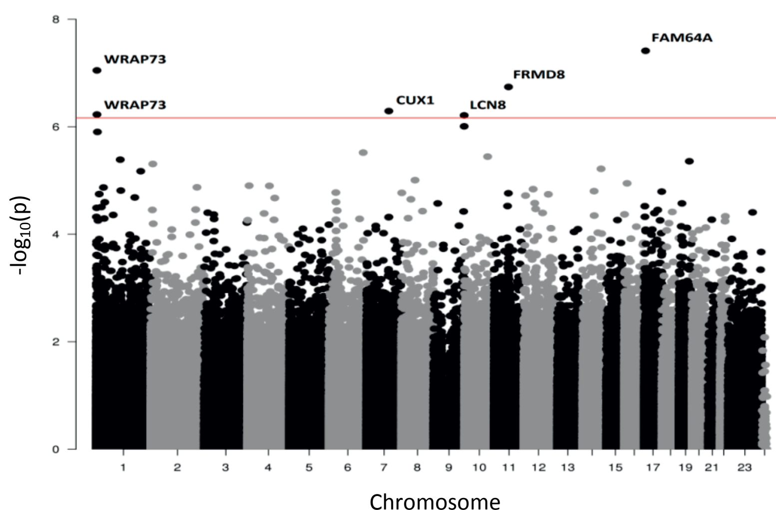


Figure 2a: Manhattan plot of the folate categorical model, adjusted for BMI, WBC counts, smoking status, physical activity, B-vitamin supplement intake, and alcohol and coffee consumption and batch effects: showing the association between folate intake and genome-wide DNA methylation (model 2), with 6 significant DMPs at FDR <0.05 (red line), in 3,894 individuals. Nearest genes for these 6 DMPs are reported.

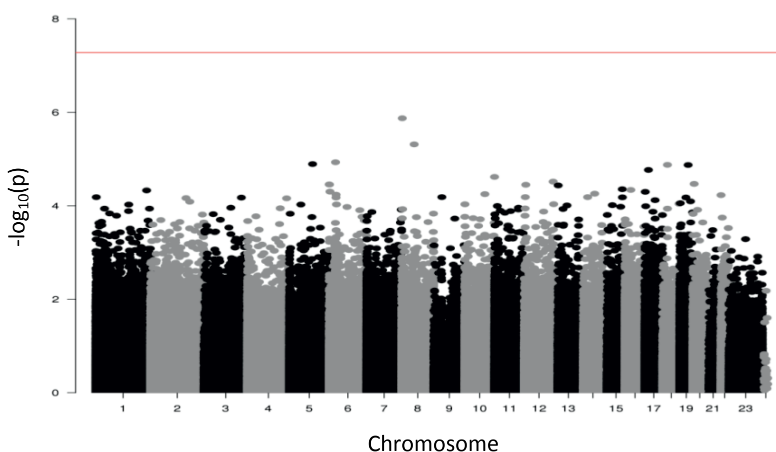


Figure 2b: Manhattan plot of the vitamin B12 categorical model, adjusted for BMI, WBC counts, smoking status, physical activity, B-vitamin supplement intake, and alcohol and coffee consumption and batch effects: showing the association between vitamin B12 intake and genome-wide DNA methylation (model 2), with no significant DMPs at FDR <0.05 (red line), in 3,566 individuals.

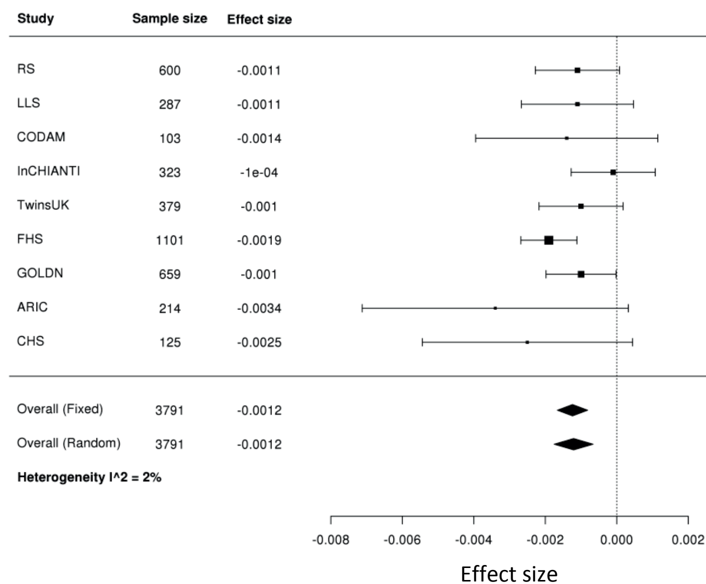


Figure 3a: Forest plot of the categorical folate intake model 2: showing the association between folate intake and the most significant DMP cg23465990 (*FAM64A*) across all studies and in meta-analysis of 3,791 individuals.

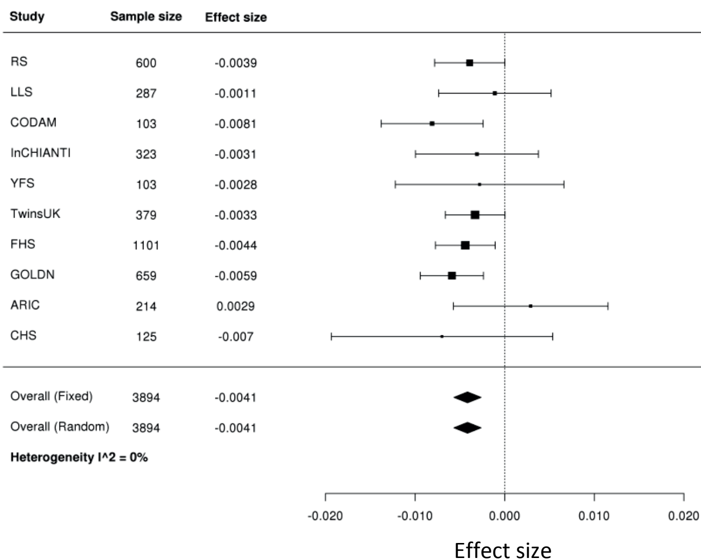


Figure 3b: Forest plot of the categorical folate intake model 2: showing the association between folate intake and the DMP cg11832534 (*WRAP73*) across all studies and in meta-analysis of 3,894 individuals.

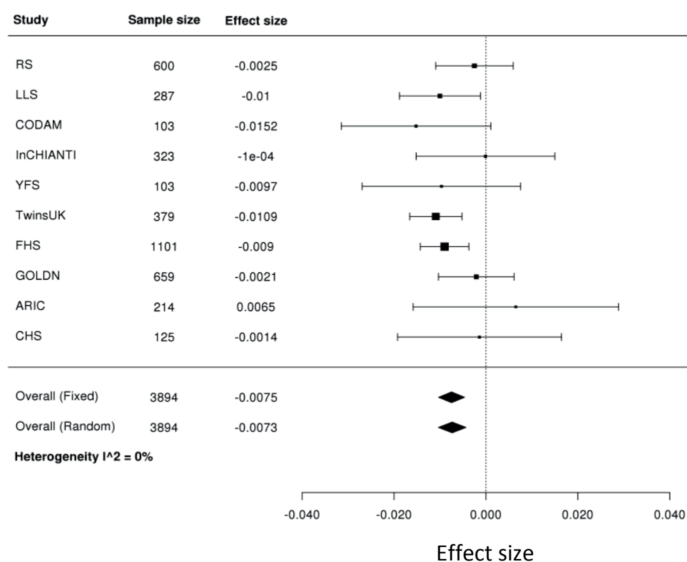


Figure 3c: Forest plot of the categorical folate intake model 2: showing the association between folate intake and the DMP cg03249011 (*FRMD8*) across all studies and in meta-analysis of 3,894 individuals.

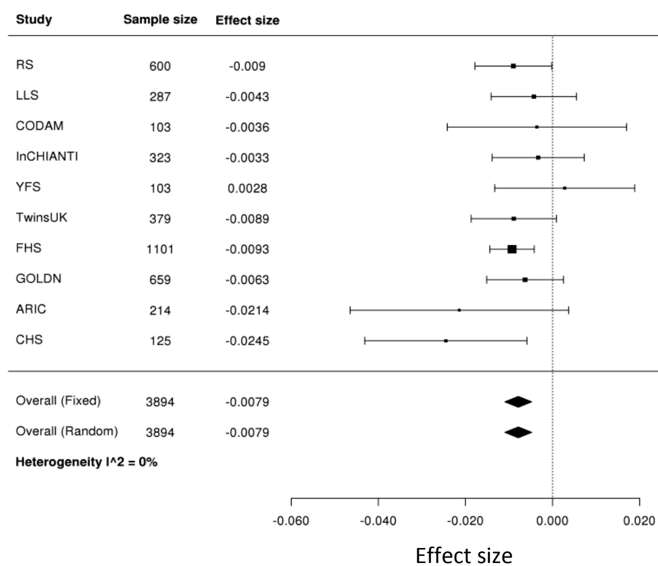


Figure 3d: Forest plot of the categorical folate intake model 2: showing the association between folate intake and the DMP cg14398883 (*CUX1*) across all studies and in meta-analysis of 3,894 individuals.

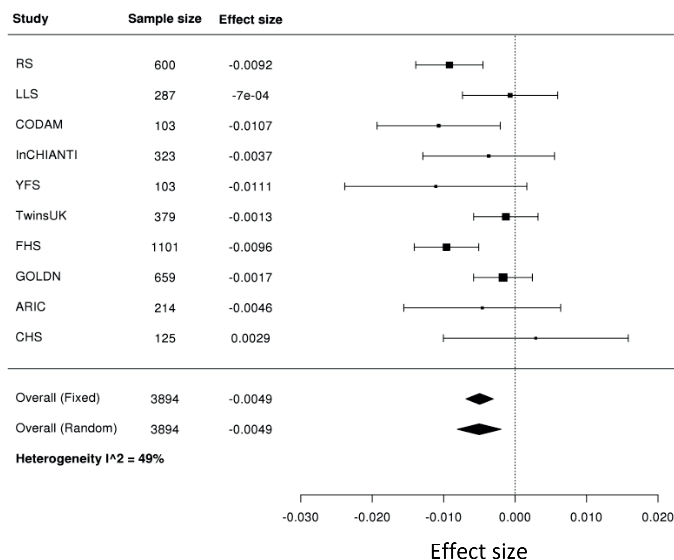


Figure 3e: Forest plot of the categorical folate intake model 2: showing the association between folate intake and the DMP cg00826902 (*WRAP73*) across all studies and in meta-analysis of 3,894 individuals.

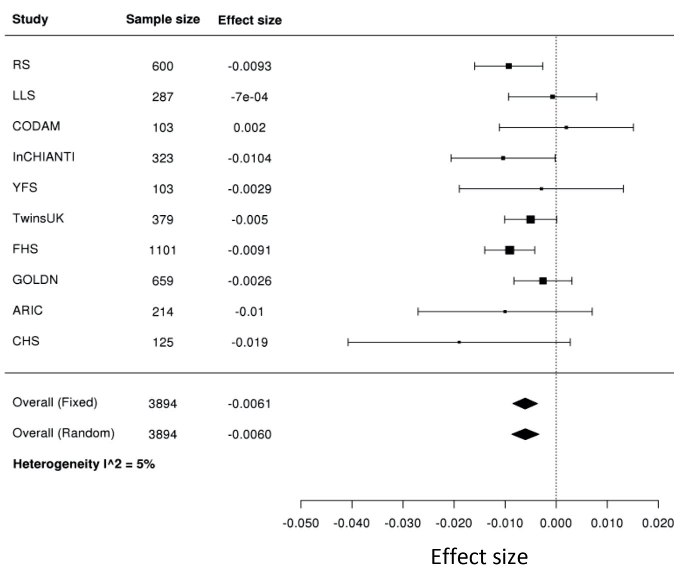


Figure 3f: Forest plot of the categorical folate intake model 2: showing the association between folate intake and the DMP cg14145338 (*LCN8*) across all studies and in meta-analysis of 3,894 individuals.

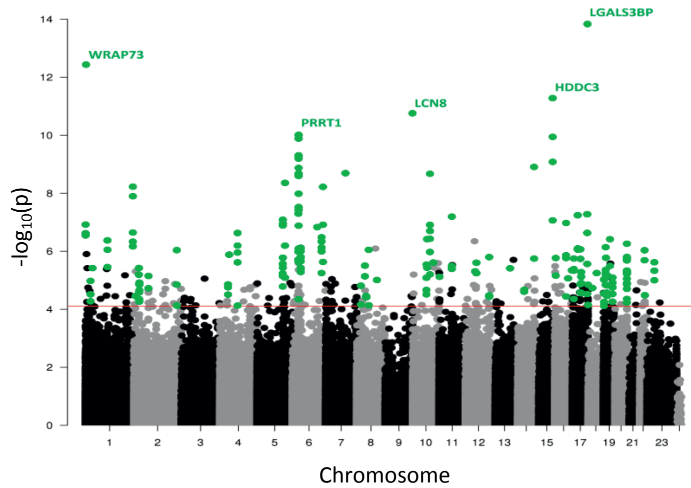


Figure 4a: Manhattan plot of the folate intake categorical model 3: showing the association between folate intake and genome-wide DNA methylation, with 74 significant DMRs at Sidak <0.05 in 3,894 samples. The red line represents the threshold of FDR <0.05 of the correlation adjusted p-values. Green circles are the CpGs that are present within the DMRs, whereas black/grey circles are the single CpGs that do not represent DMRs. Nearest genes for the top 5 DMRs are reported.

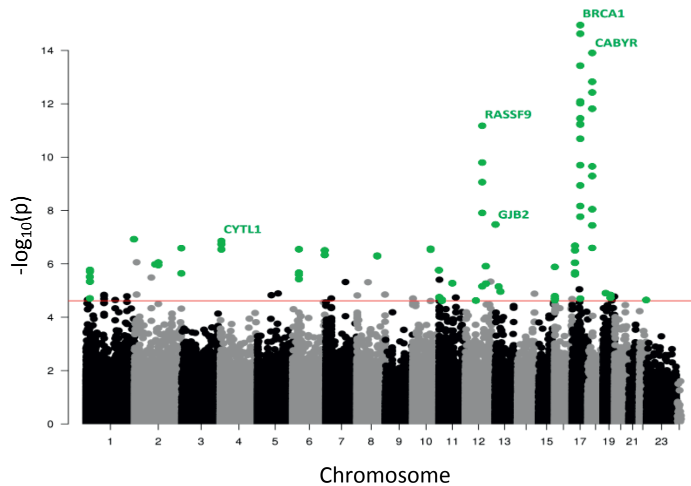


Figure 4b: Manhattan plot of the vitamin B12 intake categorical model 3: showing the association between vitamin B12 intake and genome-wide DNA methylation, with 29 significant DMRs at Sidak <0.05 in 3,565 samples. The red line represents the threshold of FDR <0.05 of the correlation adjusted p-values. Green circles are the CpGs that are present within the DMRs, whereas black/grey circles are the single CpGs that do not represent DMRs. Nearest genes for the top 5 DMRs are reported.

DISCUSSION

We have conducted the first large-scale EWAS of the association between folate and vitamin B12 intake and genome-wide DNA methylation in humans. We identified 6 novel DMPs and 74 DMRs significantly associated with dietary folate intake, and 29 DMRs significantly associated with dietary vitamin B12 intake. These novel epigenetic loci might become promising future biomarkers in folate- and vitamin B12 related pathologies.

We identified 6 novel DMPs for folate intake, of which several were located close to genes with functional relevance to centrosome, cell proliferation and tumorigenesis. The most significant DMP at cg23465990 is annotated to the gene family with sequence similarity 64 member A (*FAM64A*) on chromosome 17. FAM64A protein interacts with Calmodulin (CALM) and leukemogenic CALM/AF10 fusion protein, where it is identified as a marker for cell proliferation control [44, 45]. The DMPs at cg11832534 and cg00826902 on chromosome 1 are annotated to WD repeat containing, antisense to TP73 (*WRAP73*). WRAP73 protein is localized in the centrosome and it ensures proper spindle morphology [46]. It is a centriolar satellite and functions in ciliogenesis by promoting ciliary vesicle docking [47]. The DMP at cg14398883 on chromosome 7 is annotated to cut like homeobox 1 (*CUX1*). CUX1 is a tumor suppressor and its deficiency leads to tumor growth via the activity of phosphoinositide 3-kinase signaling [48]. The DMP at cg03249011 annotated to FERM domain containing 8 (*FRMD8*) is in an enhancer region on chromosome 11. *FRMD8* is associated with survival rate in patients of lung adenocarcinoma [49]. Lastly, the DMP at cg14145338 is annotated to lipocalin 8 (*LCN8*) on chromosome 9. *LCN8*, along with other related lipocalin genes, express in epididymis and are suggested to be involved in male fertility [50]. Further studies are needed to understand how these CpGs with seemingly no obvious link with B-vitamin homeostasis are involved in folate metabolism. In addition, validation is needed to characterize the causal relation of vitamin B intake on the function of these CpGs on nearby genes and one-carbon metabolism.

In contrast to the relatively small number of DMPs for folate intake, we observed 74 significant DMRs in association with folate intake, with several of potential relevance to immune function and stem cell function. The most significant DMR was the chr17 locus spanning 414 bp region including 8 CpGs annotating to the galectin 3 binding protein (*LGALS3BP*) gene, which is implicated in immune response. Signaling pathways regulating pluripotency of stem cells are overrepresented among the genes annotated near the folate associated DMR CpGs, comprised of AKT serine/threonine kinase 3 (*AKT3*), inhibitor of DNA binding 2, HLH protein (*ID2*), Wnt family member 6 (*WNT6*), *WNT9B* and *WNT10A* genes. This suggests a relationship between folate intake and methylation of genes involved in development related pathways [51]. In addition, downregulation of *LGALS3BP* associates with some cancers through WNT

signalling [52, 53]. Studies show a relation between low folate intake and some cancers [54-57], indicating that this relation could be via hypermethylation of *LGALS3BP*.

We identified one significant DMP at cg18642459 on chromosome 12 annotated to the *FGF6* gene. FGF6 plays a role in cellular differentiation and a potential role in muscle regeneration [58-60]. Vitamin B12 was previously shown to play a role in reparative regeneration of skeletal muscle [61] which could involve hypermethylation of the *FGF6* gene. In contrast to the limited findings for vitamin B12 and DMP, regional analysis showed 29 significant DMRs associated with vitamin B12 intake. The most significant DMR was the chr18 locus spanning 1,111 bp containing 18 DMPs and a promoter of calcium binding tyrosine phosphorylation regulated (*CABYR*). CABYR is expressed in lung and hepatocellular cancers [62, 63]. No significant pathways were observed in the enrichment analysis for the vitamin B12 associated DMR genes, possibly due to the limited number of genes. Differential methylation findings for the vitamin B12, were similar to those of folate, in that DMR associations were more numerous than those of DMP. Surprisingly, there was no overlap of DMPs or DMRs between the folate and vitamin B12 intake despite the correlation between folate and vitamin B12 intake. Thus, suggesting that their roles may be specific to different genes and pathways.

To investigate whether the previously identified transgenerational effects of maternal folate intake or levels on methylation in newborns, also generalize in adults [19, 20], we conducted a look up in our population-based meta-analysis of leukocytes in adults. A minority showed nominal significance with mostly negative direction of effects, and no significant enrichment was found in these previously identified folate-associated DMPs in both folate and vitamin B12 intake in our study. This suggests that the folate-related DMPs in newborns and adults are not similar and might differ across the life course.

In a two-year intervention study (BPROOF) with folic acid and vitamin B12, a single DMP (cg19380919) annotated to endothelial PAS domain protein 1 (*EPAS1*) was reported compared to placebo. This DMP was not significant in our meta-analyzed findings for either folate ($p = 0.78$) or vitamin B12 intake ($p = 0.43$) [Table S6]. A possible explanation for this difference is that the intervention study was specific to individuals with elevated homocysteine (greater than 15 µmol/L) [43], whereas participants in our study were generally healthy. In addition, in this intervention study the change in methylation over a 2-year time period was assessed, which might lead to different findings compared to our association study in which data at a single time point was investigated.

We hypothesized that low levels of folate give rise to DNA hypomethylation [13, 16, 17]. However, all 6 folate associated DMPs and most (73/74) DMRs were negatively associated with folate intake. For vitamin B12 intake, we see this negative association in 15 of the 29 DMRs. The occurrence of general hypomethylation with

higher folate is contrary to our hypothesis, where higher folate would transfer additional methyl groups and result in relative genome-wide hypermethylation. However, our results are in line with the study from Ono et al., who showed that higher folate intake was associated with lower global methylation [15]. Our results are also in line with the studies with maternal folate intake which showed a majority of identified DMPs in newborns to be negatively associated to folate intake [19, 20]. An intervention study of folic acid supplementation in a mice demonstrated inhibition of methylenetetrahydrofolate reductase (MTHFR) activity and reduced S-adenosylmethionine (SAM) and SAM/S-adenosylhomocysteine [64]. The inhibition of MTHFR due to higher folate intake may explain the relatively lower DNA methylation observed in our study.

Our study is also consistent with earlier work in that we observed more DMPs and DMRs correlated with folate intake as compared to vitamin B12 intake [13-17]. Based on the amount of DMPs and DMRs identified, folate seems to have a larger effect on DNA methylation. This may be due to impaired vitamin B12 uptake by medications, health disorders, genetic defects and other lifestyle factors. As a result of these conditions, vitamin B12 intake in elderly people does not necessarily reflect vitamin B12 status [65]. Alternately, folate may have a larger relationship with leukocyte DNA methylation as compared to vitamin B12 [13-17].

The strengths of our study are that this is the first epigenome-wide association study and meta-analysis conducted in a large sample size of 5,841 individuals from 10 well characterized cohorts, using the Illumina 450k methylation data. The nutrient data was harmonized across studies and all studies ran similar models with the same covariates.

While our study has yielded a number of interesting findings, among the most important insights is that folate or vitamin B12 intake, a major determinant of B-vitamin status, is not related to large-scale differences in genome-wide methylation profiles. There are a number of possible explanations for these observations. First, dietary data is prone to measurement errors, which can lead to misclassification and can compromise our ability to detect statistically significant associations. However, we addressed these limitations by using categorical models. While this analysis reduces statistical power due to the one-third reduction in sample size, greater effects were observed because the comparison is made between extreme tertile groups of the nutrient intakes. Second, vitamin B12 intake does not provide a good reflection of vitamin B12 status in older individuals as a consequence of issues related to absorption [66]. Given the fact that the mean ages in the majority of the cohorts were over 60 years, it is highly likely that use of intake could result in misclassification of vitamin B12 status. Third, although we removed individuals with prevalent cancer because of their potentially adopted dietary patterns [21, 22], the possibility of changes in dietary habits due to other reasons such as lifestyle changes cannot be excluded. Also, there

may be ethnic variation, since the ARIC sample is of African-American origin. Furthermore, there were differences in nutrient intake ranges across cohorts which could be related to the number and type of food items used in the FFQs, and the types of food composition reference tables used to calculate the nutrient data. Large differences in median intakes were also seen in Netherlands based studies, where the Rotterdam Study (RS) used 389 food items and the Leiden Longevity Study (LLS) used 183 food items. Fourth, despite our inclusion of the known confounders as covariates, the possibility of inter-correlations of these confounders with each other or belonging along the causal pathway could nullify the findings. Fifth, we did not account for the most common *MTHFR* 677C>T variant that affects DNA methylation through its interact with folate status [67]. Sixth, our study includes DMPs that approach borderline significance and have modest effect sizes. In such cases, the comb-p package for DMR analysis may return false positive DMRs [68]. Therefore, replication of our findings would be necessary before definite conclusions can be drawn. We should therefore interpret all findings of the DMR analysis with caution. Seventh, the variation in intake may not be large enough for overall hypomethylation. It may be that the genes and regions that are identified here, are only the most sensitive ones. Also, none of the cohorts had median intake of vitamin B12 that was lower than the European or American RDI. This could mean that the vitamin B12 intake was consistently high enough and our population was relatively healthy to prevent any measurable effects on methylation. Lastly, we acknowledge the possibility that blood cells may not be the ideal tissue for evaluating the association between B-vitamin intake and methylation, and that larger tissue-specific effects may be present but remain undetected in our study.

Among the European cohorts, the median intake of dietary folate ranged from 193 µg/d in the Leiden Longevity Study to 415 µg/d in the Rotterdam Study. This is slightly lower than the European recommended dietary allowance (RDA) of 200 µg/d [69]. The Rotterdam Study had the highest median folate intake being 415 µg/d. Among the American cohorts, the median intakes of dietary folate were 234 and 360 µg/d in the Genetics of Lipid Lowering Drugs and Diet Network (GOLDN) and CHS cohorts, respectively, which is less than the American RDA of 400 µg/d (except for ARIC) [69]. The Framingham Heart Study (FHS) and ARIC fulfilled the American RDA for their median dietary folate intake being 406 and 448 µg/d, respectively. Furthermore, there could be misclassification of folate intake in the GOLDN study. This is because dietary data was collected between 2002-2004 after the mandated folic acid fortification was implemented. However, the food composition table that was used for nutrient calculations originated from before fortification was implemented (i.e. before 1996, table 1].

CONCLUSIONS

We observed 6 DMPs associated with dietary folate intake. Regional associations showed 74 DMRs for dietary folate intake and 29 DMRs for dietary vitamin B12 intake. Our meta-analysis identified several novel differentially methylated loci that could serve as potential biomarkers in folate and vitamin B12 related pathologies. Further studies should validate results by measuring folate and vitamin B12 levels in plasma.

ADDITIONAL INFORMATION

Supplementary material is available on request.

REFERENCES

1. Friso S, Udali S, De Santis D, Choi SW. One-carbon metabolism and epigenetics. Mol Aspects Med 2017;54:28-36.
2. Kirke PN, Molloy AM, Daly LE, Burke H, Weir DG, Scott JM. Maternal plasma folate and vitamin B12 are independent risk factors for neural tube defects. Q J Med 1993;86:703-8.
3. Molloy AM, Mills JL, Kirke PN, Weir DG, Scott JM. Folate status and neural tube defects. Biofactors 1999;10:291-4.
4. Molloy AM, Kirke PN, Troendle JF, Burke H, Sutton M, Brody LC, et al. Maternal vitamin B12 status and risk of neural tube defects in a population with high neural tube defect prevalence and no folic Acid fortification. Pediatrics 2009;123:917-23.
5. Ma Y, Peng D, Liu C, Huang C, Luo J. Serum high concentrations of homocysteine and low levels of folic acid and vitamin B12 are significantly correlated with the categories of coronary artery diseases. BMC Cardiovasc Disord 2017;17:37.
6. Peng Y, Dong B, Wang Z. Serum folate concentrations and all-cause, cardiovascular disease and cancer mortality: A cohort study based on 1999-2010 National Health and Nutrition Examination Survey (NHANES). Int J Cardiol 2016;219:136-42.
7. Ebesunun MO, Umahoin KO, Alonge TO, Adebusoye LA. Plasma homocysteine, B vitamins and bone mineral density in osteoporosis: a possible risk for bone fracture. Afr J Med Med Sci 2014;43:41-7.
8. Herrmann M, Peter Schmidt J, Umanskaya N, Wagner A, Taban-Shomal O, Widmann T, et al. The role of hyperhomocysteinemia as well as folate, vitamin B(6) and B(12) deficiencies in osteoporosis: a systematic review. Clin Chem Lab Med 2007;45:1621-32.
9. H.B. Castellanos-Sinco COR-P, A. Santoyo-Sánchez, J. Collazo-Jaloma, C. Martínez-Murillo, E. Montaño-Figueroa, A. Sinco-Ángeles. Megaloblastic anaemia: Folic acid and vitamin B12 metabolism. Revista Médica Del Hospital General De México 2015;78:135-143.
10. Jung AY, Botma A, Lute C, Blom HJ, Ueland PM, Kvalheim G, et al. Plasma B vitamins and LINE-1 DNA methylation in leukocytes of patients with a history of colorectal adenomas. Mol Nutr Food Res 2013;57:698-708.
11. Schernhammer ES, Giovannucci E, Kawasaki T, Rosner B, Fuchs CS, Ogino S. Dietary folate, alcohol and B vitamins in relation to LINE-1 hypomethylation in colon cancer. Gut 2010;59:794-9.
12. Piyathilake CJ, Johanning GL, Macaluso M, Whiteside M, Oelschlager DK, Heimburger DC, Grizzle WE. Localized folate and vitamin B-12 deficiency in squamous cell lung cancer is associated with global DNA hypomethylation. Nutr Cancer 2000;37:99-107.
13. Zhang FF, Morabia A, Carroll J, Gonzalez K, Fulda K, Kaur M, et al. Dietary patterns are associated with levels of global genomic DNA methylation in a cancer-free population. J Nutr 2011;141:1165-1171.
14. Perng W, Villamor E, Shroff MR, Nettleton JA, Pilsner JR, Liu Y, Diez-Roux AV. Dietary intake, plasma homocysteine, and repetitive element DNA methylation in the Multi-Ethnic Study of Atherosclerosis (MESA). Nutr Metab Cardiovasc Dis 2014;24:614-622.
15. Ono H, Iwasaki M, Kuchiba A, Kasuga Y, Yokoyama S, Onuma H, et al. Association of dietary and genetic factors related to one-carbon metabolism with global methylation level of leukocyte DNA. Cancer Sci 2012;103:2159-2164.
16. Zhang FF, Santella RM, Wolff M, Kappil MA, Markowitz SB, Morabia A. White blood cell global methylation and IL-6 promoter methylation in association with diet and lifestyle risk factors in a cancer-free population. Epigenetics 2012;7:606-14.
17. Agodi A, Barchitta M, Quattrocchi A, Maugeri A, Canto C, Marchese AE, Vinciguerra M. Low fruit consumption and folate deficiency are associated with LINE-1 hypomethylation in women of a cancer-free population. Genes Nutr 2015;10:480.

18. Amarasekera M, Martino D, Ashley S, Harb H, Kesper D, Strickland D, et al. Genome-wide DNA methylation profiling identifies a folate-sensitive region of differential methylation upstream of ZFP57-imprinting regulator in humans. FASEB J 2014;28:4068-4076.
19. Gonseth S, Roy R, Houseman EA, de Smith AJ, Zhou M, Lee ST, et al. Periconceptional folate consumption is associated with neonatal DNA methylation modifications in neural crest regulatory and cancer development genes. Epigenetics 2015;10:1166-1176.
20. Joubert BR, den Dekker HT, Felix JF, Bohlin J, Ligthart S, Beckett E, et al. Maternal plasma folate impacts differential DNA methylation in an epigenome-wide meta-analysis of newborns. Nat Commun 2016;7:10577.
21. Supportive PDQ, Palliative Care Editorial B. Nutrition in Cancer Care (PDQ(R)): Health Professional Version. 2002.
22. Hebuterne X, Lemarie E, Michallet M, de Montreuil CB, Schneider SM, Goldwasser F. Prevalence of malnutrition and current use of nutrition support in patients with cancer. JPEN J Parenter Enteral Nutr 2014;38:196-204.
23. Hanley MP, Hahn MA, Li AX, Wu X, Lin J, Wang J, et al. Genome-wide DNA methylation profiling reveals cancer-associated changes within early colonic neoplasia. Oncogene 2017;36:5035-5044.
24. Ye D, Jiang D, Li Y, Jin M, Chen K. The role of LINE-1 methylation in predicting survival among colorectal cancer patients: a meta-analysis. Int J Clin Oncol 2017;22:749-757.
25. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, et al. High density DNA methylation array with single CpG site resolution. Genomics 2011;98:288-95.
26. Lehne B, Drong AW, Loh M, Zhang W, Scott WR, Tan ST, et al. A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. Genome Biol 2015;16:37.
27. Pidsley R, CC YW, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. BMC Genomics 2013;14:293.
28. Maksimovic J, Gordon L, Oshlack A. SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. Genome Biol 2012;13:R44.
29. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, Beck S. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. Bioinformatics 2013;29:189-96.
30. Fortin JP, Labbe A, Lemire M, Zanke BW, Hudson TJ, Fertig EJ, et al. Functional normalization of 450k methylation array data improves replication in large cancer studies. Genome Biol 2014;15:503.
31. Willett WC, Howe GR, Kushi LH. Adjustment for total energy intake in epidemiologic studies. Am J Clin Nutr 1997;65:1220S-1228S; discussion 1229S-1231S.
32. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics 2012;13:86.
33. Preacher KJ, Rucker DD, MacCallum RC, Nicewander WA. Use of the extreme groups approach: a critical reexamination and new recommendations. Psychol Methods 2005;10:178-92.
34. Magi R, Morris AP. GWAMA: software for genome-wide association meta-analysis. BMC Bioinformatics 2010;11:288.
35. Bonder MJ, Luijk R, Zhernakova DV, Moed M, Deelen P, Vermaat M, et al. Disease variants alter transcription factor levels and methylation of their binding sites. Nat Genet 2017;49:131-138.
36. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing Journal of the Royal Statistical Society 1995;57:289-300.

37. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, et al. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol 2010;28:495-501.
38. Speir ML, Zweig AS, Rosenbloom KR, Raney BJ, Paten B, Nejad P, et al. The UCSC Genome Browser database: 2016 update. Nucleic Acids Res 2016;44:D717-25.
39. Devlin B, Roeder K. Genomic control for association studies. Biometrics 1999;55:997-1004.
40. Liu L, Li Y, Tollefsbol TO. Gene-environment interactions and epigenetic basis of human diseases. Curr Issues Mol Biol 2008;10:25-36.
41. Pedersen BS, Schwartz DA, Yang IV, Kechris KJ. Comb-p: software for combining, analyzing, grouping and correcting spatially correlated P-values. Bioinformatics 2012;28:2986-8.
42. Wang J, Duncan D, Shi Z, Zhang B. WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. Nucleic Acids Res 2013;41:W77-83.
43. Kok DE, Dhonukshe-Rutten RA, Lute C, Heil SG, Uitterlinden AG, van der Velde N, et al. The effects of long-term daily folic acid and vitamin B12 supplementation on genome-wide DNA methylation in elderly subjects. Clin Epigenetics 2015;7:121.
44. Archangelo LF, Glasner J, Krause A, Bohlander SK. The novel CALM interactor CATS influences the subcellular localization of the leukemogenic fusion protein CALM/AF10. Oncogene 2006;25:4099-109.
45. Archangelo LF, Greif PA, Holzel M, Harasim T, Kremmer E, Przemeck GK, et al. The CALM and CALM/AF10 interactor CATS is a marker for proliferation. Mol Oncol 2008;2:356-67.
46. Hori A, Morand A, Ikebe C, Frith D, Snijders AP, Toda T. The conserved Wdr8-hMsd1/SSX2IP complex localises to the centrosome and ensures proper spindle length and orientation. Biochem Biophys Res Commun 2015;468:39-45.
47. Kurtulmus B, Wang W, Ruppert T, Neuner A, Cerikan B, Viol L, et al. WDR8 is a centriolar satellite and centriole-associated protein that promotes ciliary vesicle docking during ciliogenesis. J Cell Sci 2016;129:621-36.
48. Wong CC, Martincorena I, Rust AG, Rashid M, Alifrangis C, Alexandrov LB, et al. Inactivating CUX1 mutations promote tumorigenesis. Nat Genet 2014;46:33-8.
49. Galvan A, Frullanti E, Anderlini M, Manenti G, Noci S, Dugo M, et al. Gene expression signature of non-involved lung tissue associated with survival in lung adenocarcinoma patients. Carcinogenesis 2013;34:2767-73.
50. Suzuki K, Lareyre JJ, Sanchez D, Gutierrez G, Araki Y, Matusik RJ, Orgebin-Crist MC. Molecular evolution of epididymal lipocalin genes localized on mouse chromosome 2. Gene 2004;339:49-59.
51. Okita K, Yamanaka S. Intracellular signaling pathways regulating pluripotency of embryonic stem cells. Curr Stem Cell Res Ther 2006;1:103-11.
52. Piccolo E, Tinari N, D'Addario D, Rossi C, Iacobelli V, La Sorda R, et al. Prognostic relevance of LGALS3BP in human colorectal carcinoma. J Transl Med 2015;13:248.
53. Lee JH, Bae JA, Lee JH, Seo YW, Kho DH, Sun EG, et al. Glycoprotein 90K, downregulated in advanced colorectal cancer tissues, interacts with CD9/CD82 and suppresses the Wnt/beta-catenin signal via ISGylation of beta-catenin. Gut 2010;59:907-17.
54. Zhao Y, Guo C, Hu H, Zheng L, Ma J, Jiang L, et al. Folate intake, serum folate levels and esophageal cancer risk: an overall and dose-response meta-analysis. Oncotarget 2017;8:10458-10469.
55. Du L, Wang Y, Zhang H, Zhang H, Gao Y. Folate intake and the risk of endometrial cancer: A meta-analysis. Oncotarget 2016;7:85176-85184.
56. Donnenfeld M, Deschasaux M, Latino-Martel P, Diallo A, Galan P, Hercberg S, et al. Prospective association between dietary folate intake and skin cancer risk: results from

the Supplementation en Vitamines et Mineraux Antioxydants cohort. Am J Clin Nutr 2015;102:471-8.

57. He H, Shui B. Folate intake and risk of bladder cancer: a meta-analysis of epidemiological studies. Int J Food Sci Nutr 2014;65:286-92.
58. Floss T, Arnold HH, Braun T. A role for FGF-6 in skeletal muscle regeneration. Genes Dev 1997;11:2040-51.
59. Armand AS, Launay T, Pariset C, Della Gaspera B, Charbonnier F, Chanoine C. Injection of FGF6 accelerates regeneration of the soleus muscle in adult mice. Biochim Biophys Acta 2003;1642:97-105.
60. Fiore F, Sebille A, Birnbaum D. Skeletal muscle regeneration is not impaired in Fgf6 -/- mutant mice. Biochem Biophys Res Commun 2000;272:138-43.
61. Kovalenko TM. Reparative regeneration of skeletal muscle in mammals receiving vitamin B12. Bull Exp Biol Med 1960;49:517-520.
62. Luo C, Xiao X, Liu D, Chen S, Li M, Xu A, et al. CABYR is a novel cancer-testis antigen in lung cancer. Clin Cancer Res 2007;13:1288-97.
63. Li H, Fang L, Xiao X, Shen L. The expression and effects the CABYR-c transcript of CABYR gene in hepatocellular carcinoma. Bull Cancer 2012;99:E26-33.
64. Christensen KE, Mikael LG, Leung KY, Levesque N, Deng L, Wu Q, et al. High folic acid consumption leads to pseudo-MTHFR deficiency, altered lipid metabolism, and liver injury in mice. Am J Clin Nutr 2015;101:646-58.
65. Institute of Medicine Standing Committee on the Scientific Evaluation of Dietary Reference I, its Panel on Folate OBV, Choline. 1998.
66. Hughes CF, Ward M, Hoey L, McNulty H. Vitamin B12 and ageing: current issues and interaction with folate. Ann Clin Biochem 2013;50:315-29.
67. Friso S, Choi SW, Girelli D, Mason JB, Dolnikowski GG, Bagley PJ, et al. A common mutation in the 5,10-methylenetetrahydrofolate reductase gene affects genomic DNA methylation through an interaction with folate status. Proc Natl Acad Sci U S A 2002;99:5606-11.
68. Kolde R, Martens K, Lokk K, Laur S, Vilo J. seqlm: an MDL based method for identifying differentially methylated regions in high density methylation array data. Bioinformatics 2016;32:2604-10.
69. Bolzetta F, Veronese N, De Rui M, Berton L, Toffanello ED, Carraro S, et al. Are the Recommended Dietary Allowances for Vitamins Appropriate for Elderly People? J Acad Nutr Diet 2015;115:1789-97.

CHAPTER 8

Epigenetic Signatures of Cigarette Smoking

Roby Joehanes*, Allan C. Just*, Riccardo E. Marioni*, Luke C. Pilling*, Lindsay M. Reynolds*, Pooja R. Mandaviya*, Weihua Guan*, Tao Xu*, Cathy E. Elks*, Stella Aslibekyan*, Hortensia Moreno-Macias*, Jennifer A. Smith*, Jennifer A Brody*; Radhika Dhingra*; Paul Yousefi, James S. Pankow, Sonja Kunze, Sonia Shah, Allan F. McRae, Kurt Lohman, Jin Sha, Devin M. Absher, Luigi Ferrucci, Wei Zhao, Ellen W. Demerath, Jan Bressler, Megan L. Grove, Tianxiao Huan, Chunyu Liu, Michael M. Mendelson, Chen Yao, Douglas P. Kiel, Annette Peters, Rui Wang-Sattler, Peter M. Visscher, Naomi R. Wray, John M. Starr, Jingzhong Ding, Carlos J. Rodriguez, Nicholas J. Wareham, Marguerite R. Irvin, Degui Zhi, Myrto Barrdahl, Paolo Vineis, André G. Uitterlinden, Albert Hofman, Joel Schwartz, Elena Colicino, Lifang Hou, MD, Pantel S. Vokonas, Dena G. Hernandez, Andrew B. Singleton, Stefania Bandinelli, Stephen T. Turner, Erin B. Ware, Alicia K. Smith, Torsten Klengel, Elisabeth B. Binder, Bruce M. Psaty, Kent D. Taylor, Sina A. Gharib, Brenton R. Swenson, Liming Liang, Dawn L. DeMeo, MD, MPH52; George T. O'Connor, Zdenko Herceg, Kerry J. Ressler, Karen N. Conneely**, Nona Sotoodehnia**, Sharon L. R. Kardia**, David Melzer**, Andrea A. Baccarelli**, Joyce B. J. van Meurs**, Isabelle Romieu**, Donna K. Arnett**, Ken K. Ong**, Yongmei Liu**, Melanie Waldenberger**, Ian J. Deary**, Myriam Fornage**, Daniel Levy**, Stephanie J. London**

**These authors contributed equally as first authors*
***These authors contributed equally as senior authors*

Circ Cardiovasc Genet. 2016 Oct;9(5):436-447.

ABSTRACT

Background - DNA methylation leaves a long-term signature of smoking exposure and is one potential mechanism by which tobacco exposure predisposes to adverse health outcomes, such as cancers, osteoporosis, lung, and cardiovascular disorders.

Methods and Results - To comprehensively determine the association between cigarette smoking and DNA methylation, we conducted a meta-analysis of genome-wide DNA methylation assessed using the Illumina BeadChip 450K array on 15,907 blood derived DNA samples from participants in 16 cohorts (including 2,433 current, 6,518 former, and 6,956 never smokers). Comparing current versus never smokers, 2,623 CpG sites (CpGs), annotated to 1,405 genes, were statistically significantly differentially methylated at Bonferroni threshold of $p<1x10^{-7}$ (18,760 CpGs at False Discovery Rate (FDR)<0.05). Genes annotated to these CpGs were enriched for associations with several smoking-related traits in genome-wide studies including pulmonary function, cancers, inflammatory diseases and heart disease. Comparing former versus never smokers, 185 of the CpGs that differed between current and never smokers were significant p<1x10-7 (2,623 CpGs at FDR<0.05), indicating a pattern of persistent altered methylation, with attenuation, after smoking cessation. Transcriptomic integration identified effects on gene expression at many differentially methylated CpGs.

Conclusions - Cigarette smoking has a broad impact on genome-wide methylation that, at many loci, persists many years after smoking cessation. Many of the differentially methylated genes were novel genes with respect to biologic effects of smoking, and might represent therapeutic targets for prevention or treatment of tobacco-related diseases. Methylation at these sites could also serve as sensitive and stable biomarkers of lifetime exposure to tobacco smoke.

INTRODUCTION

Cigarette smoking is a major causal risk factor for various diseases including cancers, cardiovascular disease (CVD), chronic obstructive pulmonary disease (COPD)[1], and osteoporosis[1]. Worldwide cessation campaigns and legislative actions have been accompanied by a reduction in the number of cigarette smokers and corresponding increases in the number of former smokers. In the US, there are more former smokers than current smokers[1]. Despite the decline in the prevalence of smoking in many countries, it remains the leading preventable cause of death in the world, accounting for nearly 6 million deaths each year[2].

Even decades after cessation, cigarette smoking confers long-term risk of diseases including some cancers, chronic obstructive pulmonary disease, and stroke[1]. The mechanisms for these long-term effects are not well understood. DNA methylation changes have been proposed as one possible explanation.

DNA methylation appears to reflect exposure to a variety of lifestyle factors[3], including cigarette smoking. Several studies have shown reproducible associations between tobacco smoking and altered DNA methylation at multiple cytosine-phosphate-guanine (CpG) sites[4–15]. Some DNA methylation sites associated with tobacco smoking have also localized to genes related to coronary heart disease[5] and pulmonary disease[16]. Some studies have found different associated CpGs in smokers versus non-smokers[8,11]. Consortium-based meta-analyses have been extremely successful in identifying genetic variants associated with numerous phenotypes, but large-scale meta-analyses of genome-wide DNA methylation data have not yet been widely employed. It is likely that additional novel loci, differentially methylated in response to cigarette smoking remain to be discovered by meta-analyzing data across larger sample sizes comprising multiple cohorts. Differentially methylated loci with respect to smoking may serve as biomarkers of lifetime smoking exposure. They may also shed light on the molecular mechanisms by which tobacco exposure predisposes to multiple diseases.

A recent systematic review[13] analyzed published findings across 14 epigenome-wide association studies of smoking exposure across various DNA methylation platforms of varying degrees of coverage and varying phenotypic definitions. Among these were 12 studies (comprising 4,750 subjects) that used the more comprehensive Illumina Human Methylation BeadChip 450K array (Illumina 450K), which includes and greatly expands on the coverage of the earlier 27K platform. The review compares only statistically significant published results and is not a meta-analysis which can identify signals that do not reach statistical significance in individual studies[17].

In the current study, we meta-analyzed association results between DNA methylation and cigarette smoking in 15,907 individuals from 16 cohorts in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium

using a harmonized analysis. Methylation was measured on DNA extracted from blood samples using the Illumina Human Methylation BeadChip 450K array. In separate analyses, we compared current smokers and past smokers to non-smokers and characterized the persistence of smoking-related CpG methylation associations with the duration of smoking cessation among former smokers. We integrated information from genome-wide association studies (GWAS) and gene expression data to gain insight into potential functional relevance of our findings for human diseases. Finally we conducted analyses to identify pathways that may explain the molecular effects of cigarette exposure on tobacco-related diseases.

MATERIALS AND METHODS

Study participants

This study comprised a total of 15,907 participants from 16 cohorts of the Cohorts for Heart and Aging Research in Genetic Epidemiology Consortium (Supplementary Table 1). The 16 participating cohorts are ARIC, FHS Offspring, KORA F4, GOLDN, LBC 1921, LBC 1936, NAS, Rotterdam, Inchianti, GTP, CHS European Ancestry (EA), CHS African Ancestry (AA), GENOA, EPIC Norfolk, EPIC, and MESA. Of these, 12,161 are of European Ancestry (EA) and 3,746 are of African Ancestry (AA). The study was approved by institutional review committees for each cohort and all participants provided written informed consent for genetic research.

DNA methylation sample and measurement

For most studies, methylation was measured on DNA extracted from whole blood, but some studies used CD4+ T cells or monocytes (Supplementary Table 1). In all studies, DNA was bisulfite-converted using the Zymo EZ DNA methylation kit and assayed for methylation using the Infinium HumanMethylation 450 BeadChip, which contains 485,512 CpG sites. Details of genomic DNA preparation, bisulfite conversion, and methylation assay for each cohort can be found in the online Supplementary Materials.

Raw methylated and total probe intensities were extracted using the Illumina Genome Studio methylation module. Preprocessing of the methylated signal (M) and unmethylated signal (U) was conducted using various software tools, primarily DASEN of wateRmelon[18] and BMIQ[19], both of which are R packages. The methylation beta (β) values were defined as $\beta = M/(M+U)$. Each cohort followed its own quality control protocols, removing poor quality or outlier samples and excluding low quality CpG sites (with detection p-value>0.01). Each cohort evaluated batch effects and controlled for them in the analysis. Details of these processes can be found in the online Supplementary Materials.

Smoking phenotype definition

Self-reported cigarette smoking status was divided into three categories. Current smokers were defined as those who have smoked at least one cigarette a day within 12 months prior to the blood draw, former smokers were defined as those who had ever smoked at least one cigarette a day, but had stopped at least 12 months prior to the blood draw, and never smokers reported never having smoked. Pack years was calculated based on self-report as the average number of cigarettes per day smoked divided by 20 multiplied by the number of years of smoking, with zero assigned to never smokers. A few cohorts recorded the number of years since each former smoker had stopped smoking.

Cohort specific analyses and meta-analysis

Each cohort analyzed its data using at least two linear mixed effect models. Each model was run separately for each CpG site. Model 1 is as follows:

$$\beta = \text{Smoking phenotype} + \text{Sex} + \text{Age} + \text{blood count} + \text{technical covariates}, \quad (1)$$

where blood count comprises the fractions of CD4+ T-cells, CD8+ T-cells, NK cells, monocyte, and eosinophils either measured or estimated using the Houseman et al. method[20]. The blood count adjustment was performed only in cohorts with whole blood and leukocyte samples. Familial relationship was also accounted for in the model when applicable (*e.g.*, for FHS, see Supplementary Material for details). Acknowledging that each cohort may be influenced by a unique set of technical factors, we allow each cohort to choose its cohort-specific technical covariates. Model 2 added to model 1 body mass index (BMI) because it is associated with methylation at some loci, making it a potential confounder[21]. Only three cohorts participated in model 2 analysis: FHS, KORA, and NAS. Model 3 substituted smoking phenotypes for pack years. Only three cohorts participated in model 3 analysis: FHS, Rotterdam, and Inchianti. The pack year analysis was performed only on two subsets: current *vs*. never smokers and former *vs*. never smokers. Combining all three categories would require accurate records of time of quitting, which among the three cohorts was available for only FHS. To investigate cell type differences, we removed blood counts from Model 1 and called it Model 4. Only three cohorts participated in this analysis: FHS, KORA, and NAS. All models were run with the lme4 package[22] in R[23], except for FHS (See Supplementary Materials for details).

Meta-analysis was performed to combine the results from all cohorts. Due to the variability of available CpG sites after quality control steps, we excluded CpG sites that were available in fewer than three cohorts. The remaining 485,381 CpG sites were then meta-analyzed with a random-effects model using the following formula:

$$E_i = \mu + s_i + e_i, \quad (2)$$

where E_i is the observed effect of study i, μ is the main smoking effect, s_i is the between-study error for study i, and e_i is the within-study error for study i, with both s_i

and e_i are assumed to be normally distributed. The model is fitted using the restricted maximum likelihood (REML) criterion in R's *metafor*[24] package. Multiple-testing adjustment on the resulting p-values was performed using the False Discovery Rate (FDR) method of Benjamini and Hochberg[25]. In addition, we also report results using the Bonferroni-corrected threshold of 1×10^{-7} (~ 0.05/485,381).

The regression coefficient β (from meta-analysis) is interpretable as the difference in mean methylation between current and never smokers. We multiplied these by 100 to represent the percentage methylation difference where methylation ranges from 0-100%.

Literature review to identify genes previously associated with smoking and methylation

We used the same literature search strategy published previously26. A broad query of NCBI's PubMed literature database using medical subject heading (MeSH) terms ("((((DNA Methylation[Mesh]) OR methylation)) AND ((Smoking[Mesh]) OR smoking))") yielded 775 results when initially performed on January 8, 2015 and 789 studies when repeated to update the results on March 1, 2015. Results were reviewed by abstract to determine whether studies met inclusion criteria: 1) performed in healthy human populations, 2) agnostically examined >1,000 CpG sites at a time, 3) only cigarette exposure was considered, and 4) with public reporting of P-values and gene annotations. A total of 25 publications met inclusion criteria, listed in the fourth supplementary table of Joubert et al.26. CpG level results (P-values and gene annotations) for sites showing genome-wide statistically significant associations (FDR<0.05) were extracted and resulted in 1,185 genes previously associated with adult or maternal smoking. All CpGs annotated to these 1,185 genes were marked as "previously found".

Gene set enrichment analysis (GSEA)

Gene-Set Enrichment Analysis (GSEA)[27] was performed in the website (http://software.broadinstitute.org/gsea/msigdb/annotate.jsp) on significant findings to determine putative functions of the CpG sites. We selected gene ontology (GO) biological process (C5-BP) and collected all categories with FDR<0.05 (up to 100 categories).

Enrichment analysis for localization to different genomic features

Enrichment analysis on genomic features were performed using the annotation file supplied by the Illumina (version 1.2, downloaded from manufacturer's website, http://support.illumina.com/array/array_kits/infinium_humanmethylation450_beadchip_kit/downloads.html), which contains information of CpG location relative to gene (*i.e.*, body, first exon, 3' UTR, 5'UTR, within 200 base pairs of Transcriptional Start Site [TSS200], and TSS1500), the relation of CpG site to a CpG island (*i.e.*, island, northern

shelf, northern shore, southern shelf, and southern shore), whether the CpG site is known to be in differentially methylated regions, and whether the CpG site is known to be an enhancer or a DNAse I Hypersensitive Site (DHS). Enrichment analysis was performed using one-sided Fisher's exact set for each feature, using R's fisher.test.

Genome-wide association study (GWAS) analysis

We intersected our results with SNPs having genome-wide association study (GWAS) p-values$\leq 5 \times 10^{-8}$ in the NHGRI GWAS catalog (accessed November 2, 2015)[28]. The catalog contained 9,777 SNPs annotated to 7,075 genes associated with 865 phenotypes at $p \leq 5 \times 10^{-8}$. To determine the genes, we looked up each significant CpG on the annotation file supplied by Illumina. Enrichment analysis was performed on a per gene basis using one-sided Fisher's exact test.

For bone mineral phenotype enrichment, we included all SNPs including terms "bone mineral density" or "osteoporosis". For cardiovascular disease (CVD), we included all SNPs containing terms "cardiovascular disease", "stroke", "coronary disease", "cardiomyopathy", or "myocardial infarction". For CVD risk factors, we included all SNPs containing terms "blood pressure", "cholesterol", "diabetes", "obesity", or "hypertension". For overall cancer enrichment, we included all SNPs containing terms "cancer", "carcinoma", or "lymphoma", while removing those pertaining to cancer treatment effects. For overall pulmonary phenotype enrichment, we included all SNPs containing terms "pulmonary disease", "pulmonary function", "emphysema", "asthma", or "airflow obstruction".

Analysis of persistence of methylation signals with time since quitting smoking among former smokers

We examined whether smoking methylation associations were attenuated over time in the FHS cohort, which had ascertained longitudinal smoking status of over 35 years. The analysis was performed on seven dichotomous variables, indicating cessation of smoking for 5, 10, 15, 20, 25, and 30 years versus never smokers. For example, for five year cessation variable, those who quit smoking five years or more are marked as ones, while never smokers are marked as zeroes and current smokers are excluded. For this analysis, we used the *pedigreemm* package[29] with the same set of covariates as in the primary analysis. Sites with $p<0.002$ across all seven variables were deemed to be statistically significant compared to never-smoker levels.

Methylation by expression (MxE) analysis

To determine transcriptomic association of each significant CpG site, we interrogated such CpG sites in the FHS gene-level methylation by expression (MxE) database, at genome-wide false discovery rate (FDR)<0.05. The MxE database was constructed from 2,262 individuals from the FHS Offspring cohort attending examination cycle eight (2005-2008) with both whole blood DN methylation and transcriptomic data based on

the Affymetrix Human Exon Array ST 1.0. Enrichment analysis was performed using a one-sided Fisher's exact test. We defined that the methylation CpG site and the corresponding transcript are associated in *cis* if the location of the CpG site is within 500 kilobases of the transcript's start location.

Analysis of ethnic discrepancy between African Ancestry (AA) and European Ancestry (EA) cohorts

Meta-analysis of the current versus never smoker results of EA cohorts (FHS, KORA, GOLDN, LBC 1921, LBC 1936, NAS, Rotterdam, Inchianti, EPIC, EPIC Norfolk, MESA, CHS-EA) was performed separately from those of AA cohorts (ARIC, GTP, GENOA, CHS-AA).

Analysis of samples types for DNA extraction

Meta-analysis was performed on the results from cohorts with whole blood/buffy coat samples (FHS, KORA, LBC 1921, LBC 1936, NAS, Rotterdam, Inchianti, GTP, CHS-EA, CHS-AA, ARIC, GENOA, EPIC, and EPIC-Norfolk). CD4+ samples in GOLDN and CD14+ samples in MESA, because they comprise single cohorts, are not meta-analyzed. Correlations of results across different cell types were performed on CpG sites with FDR<0.05 in at least one cell type.

RESULTS

Table 1 displays the characteristics of participants in the meta-analysis. The proportion of participants reporting current smoking ranged from 4% to 33% across the different study populations. The characteristics of the participants within each cohort are provided in Supplementary Table 1.

Current versus Never Smokers

In the meta-analysis of current cigarette smokers (N=2,433) versus never smokers (N=6,956), 2,623 CpGs annotated to 1,405 genes met Bonferroni significance after correction for 485,381 tests ($P<1x10^{-7}$). Based on genome-wide false discovery rate (FDR)<0.05, 18,760 CpG sites (CpGs) annotated to 7,201 genes were differentially methylated. There was a moderate inflation factor[30] λ of 1.32 (Supplementary Figure 1), which is consistent with a large number of sites being impacted by smoking. Our results lend support many previously reported loci[7,8,11,13], including CpGs annotated to *AHRR*, *RARA*, *F2RL3*, and *LRRN3* (Supplementary Table 2). Not surprisingly, cg05575921 annotated to *AHRR*, the top CpG identified in most prior studies of smoking, was highly significant in our meta-analysis ($P=4.6x10^{-26}$; ranked 36, Supplementary Table 2) and also had the largest effect size (-18% difference in methylation) which is comparable to effect sizes in previous studies[18]. Of the 18,760 significant CpGs at FDR<0.05, 16,673 (annotated to 6,720 genes) have not been previously reported to be associated with cigarette smoking – these include 1,500 of the 2,623 CpGs that met Bonferroni significance. The 25 CpGs with lowest p-values for both overall and novel findings are

shown in Table 2. Supplementary Table 2 provides the complete list of all CpGs that were significantly differentially methylated (FDR<0.05) in analysis of current versus never smokers. Adding body mass index (BMI) into the model did not appreciably alter the results (Supplementary Figure 2).

Methylation can be either reduced or increased at CpG sites in response to smoking. For the 53.2% of FDR significant CpGs with increased methylation in response to current smoking the mean percentage difference in methylation between current and never smokers was 0.5% (SD=0.37%, range 0.06-7.3%). For 46.8% of CpGs with decreased methylation in response to current smoking the mean percentage difference was 0.65% (SD=0.56, range 0.04-18%). The volcano plot can be found in Supplementary Figure 3.

We did not observe correlation between the number of significant CpGs and either the size of the gene or the number of exons or the coverage of the methylation platform. We performed a formal enrichment test for each of the 7,201 genes in regards to the length of the gene or number of exons and found only three for which associations were observed (*AHRR*, *PRRT1*, and *TNF*). However, given the robust findings for a specific CpG in *AHRR* in multiple studies in the literature[4,7,9] as well as our own, and its key role in the AHR pathway which is crucial in the response to polyaromatic hydrocarbons, such as are produced by smoking[31], it seems very unlikely that the *AHRR* findings are false positives. Likewise there is strong support in the literature for *PRRT1*[32] and *TNF*[33]. The enrichment results for methylation platform coverage also yielded the same three genes.

In a subset of three cohorts (1,827 subjects), we investigated the association of the number of pack-years smoked with the 18,760 CpGs that were differentially methylated (FDR<0.05) between current versus never smokers. Significant dose responses were observed for 11,267 CpGs (60.1%) at FDR<0.05 (Supplementary Table 3). To investigate the pathways implicated by these genes, we performed a gene-set enrichment analysis[34] on the annotated genes. The results suggested that cigarette smoking is associated with potential changes in numerous vital molecular processes, such as signal transduction (FDR=2.8 x 10^{-79}), protein metabolic processes (FDR=1.2 x 10^{-43}), and transcription pathways (FDR=8.4 x 10^{-31}). The complete list of 99 enriched molecular processes can be found in Supplementary Table 4.

Former versus Never Smokers

Meta-analysis of former (N=6,518) versus never smokers (N=6,956) restricted to the 18,760 CpG sites that were differentially methylated in current versus never smokers identified 2,568 CpGs annotated to 1,326 genes at FDR<0.05 (Supplementary Table 5). There were 185 CpGs (annotated to 149 genes) that also met Bonferroni correction (P< 0.05/18760 ~ 2.67×10^{-6}). There was no evidence of inflation[30] (λ=0.98) (Supplementary Figure 4). We also confirmed previously reported findings for CpGs annotated to *AHRR*,

RARA, and *LRRN3*[7,8,11,13]. Effect sizes of these CpGs were all weaker than in the analysis of current versus never smokers [61.2% ±15.3% weaker] for the 2,568 CpGs that remained significantly differentially methylated in former *vs.* never smokers compared with current *vs*. never smokers. Results for the top 25 CpGs are displayed in Table 3. Adding BMI to the model did not appreciably alter the results (Supplementary Figure 5). A volcano plot can be found in Supplementary Figure 6. In a subset of three cohorts (3,349 subjects), analyses using pack-years confirmed a significant dose response for 1,804 of the 2,568 CpGs (70%) annotated to 942 genes at FDR<0.05 (Supplementary Table 6).

The gene-set enrichment analysis27 in the former versus never smoker analyses on all 1,326 genes revealed enrichment for genes associated with protein metabolic processes (FDR=1.1 x 10^{-23}), RNA metabolic processes (FDR=1.4 x 10^{-17}), and transcription pathways (FDR=3.9 x 10^{-18}) (Supplementary Table 7). The gene-set enrichment analysis on the 942 genes for which the 1,804 CpGs exhibited dose responses with pack-years also revealed similar pathways to those summarized in Supplementary Table 7, except with weaker enrichment FDR values.

In 2,648 Framingham Heart Study participants with up to 30 years of prospectively collected smoking data, we examined the 2,568 CpGs that were differentially methylated in meta-analysis of former versus never smokers and explored their associations with time since smoking cessation. Methylation levels of most CpGs returned toward that of never-smokers within five years of smoking cessation. However, 36 CpGs annotated to 19 genes, including *TIAM2*, *PRRT1*, *AHRR*, *F2RL3*, *GNG12*, *LRRN3*, *APBA2*, *MACROD2*, and *PRSS23* did not return to never-smoker levels even after 30 years of smoking cessation (Figure 1, Table 4).

The EPIC studies included cancer cases plus non-cancer controls analyzed together, adjusting for cancer status. The other studies were population-based samples not selected for disease status. To evaluate residual confounding by cancer status after adjustment, we repeated the meta-analysis without the EPIC studies. The effect estimates were highly correlated: Pearson ρ = 0.99 for current versus never smoking and 0.98 for former smoking versus never.

Table 1. Participant characteristics

Characteristics	Current Smokers	Former Smokers	Never Smokers
	N=2,433	N=6,518	N=6,956
Sex (% Male)	0.463	0.556	0.317
Age (years)*	57.7 ± 7.7	64.8 ± 8.2	61.2 ± 9.7
BMI (kg/m2)*	27.3 ± 5.4	28.7 ± 5.0	28.6 ± 5.3

*weighted mean ± pooled standard deviation across cohorts

Table 2. Most statistically significant CpG sites that were associated with current vs. never smoker status

Probe ID	Chr	Location	Gene Symbol*	Coef†	S.E.	P	FDR
25 most significant CpG sites							
cg16145216	1	42385662	*HIVEP3*	0.0298	0.002	6.7×10^{-48}	3.3×10^{-42}
cg19406367	1	66999929	*SGIP1*	0.0175	0.0013	7×10^{-44}	1.7×10^{-38}
cg05603985	1	2161049	*SKI*	−0.0122	0.0009	1.8×10^{-43}	2.8×10^{-38}
cg14099685	11	47546068	*CUGBP1*	−0.0124	0.0009	1.5×10^{-42}	1.8×10^{-37}
cg12513616	5	1.77E+08		−0.0262	0.002	6.1×10^{-41}	5.9×10^{-36}
cg03792876‡	16	73243		−0.0182	0.0014	7.2×10^{-38}	5.9×10^{-33}
cg01097768	5	378854	*AHRR*	−0.0166	0.0013	6.8×10^{-35}	4.7×10^{-30}
cg26856289	1	24307516	*SFRS13A*	−0.0163	0.0013	8.6×10^{-35}	5.2×10^{-30}
cg07954423	9	1.31E+08	*FAM102A*	−0.0134	0.0011	1.2×10^{-34}	6.3×10^{-30}
cg01940273	2	2.33E+08		−0.0815	0.0067	2×10^{-34}	9.8×10^{-30}
cg01083131	16	67877413	*THAP11;CENPT*	−0.0155	0.0013	3.7×10^{-34}	1.6×10^{-29}
cg01017464	18	47018095	*SNORD58A; SNORD58B; RPL17*	−0.0172	0.0014	1.9×10^{-33}	7.6×10^{-29}
cg06121808	2	1.13E+08	*SLC20A1*	−0.0143	0.0012	2.1×10^{-32}	7.9×10^{-28}
cg10062919	17	38503802	*RARA*	−0.0128	0.0011	9.2×10^{-32}	3.2×10^{-27}
cg20066188	22	37678791	*CYTH4*	−0.0252	0.0022	1.6×10^{-31}	5.2×10^{-27}
cg04551776	5	393366	*AHRR*	−0.0244	0.0021	5.8×10^{-31}	1.8×10^{-26}
cg11152412	15	74927688	*EDC3*	−0.0077	0.0007	1.8×10^{-30}	5×10^{-26}
cg00073090	19	1265879		−0.0196	0.0017	4.2×10^{-30}	1.1×10^{-25}
cg11902777	5	368843	*AHRR*	−0.0201	0.0018	9.1×10^{-30}	2.3×10^{-25}
cg25212453	17	1509953	*SLC43A2*	−0.0101	0.0009	1.4×10^{-29}	3.5×10^{-25}
cg04956244	17	38511592	*RARA*	0.0122	0.0011	1.5×10^{-29}	3.5×10^{-25}
cg13951797	16	2204381	*TRAF7*	−0.0153	0.0014	1.6×10^{-29}	3.5×10^{-25}
cg11028075	10	97200911	*SORBS1*	0.0175	0.0016	1.7×10^{-29}	3.6×10^{-25}
cg11700584†	14	50088544	*RPL36AL;MGAT2*	−0.0151	0.0013	3.4×10^{-29}	6.8×10^{-25}
cg11263997	11	70257280	*CTTN*	0.005	0.0005	4.3×10^{-29}	8.4×10^{-25}

Probe ID	Chr	Location	Gene Symbol*	Coef†	S.E.	P	FDR
25 most significant novel CpG sites							
cg11700584	14	50088544	*RPL36AL; MGAT2*	−0.0151	0.0013	3.4×10^{-29}	6.8×10^{-25}
cg22417733	6	1.53E+08	*FBXO5*	−0.0171	0.0015	1.5×10^{-28}	2.7×10^{-24}
cg08118908	16	15787920	*NDE1*	0.0053	0.0005	5.4×10^{-26}	7.1×10^{-22}
cg14003265	9	1.4E+08	*TRAF2*	−0.0106	0.001	3.2×10^{-25}	3.7×10^{-21}
cg02556393	3	1.69E+08	*MECOM*	−0.0162	0.0016	2.8×10^{-24}	2.6×10^{-20}
cg01218206	11	1.17E+08	*SIK3*	−0.0150	0.0015	3.1×10^{-23}	2.5×10^{-19}
cg04987734	14	1.03E+08	*CDC42BPB*	0.0149	0.0015	9.0×10^{-23}	6.8×10^{-19}
cg27118035	16	31891978	*ZNF267*	0.0136	0.0014	2.4×10^{-22}	1.7×10^{-18}
cg18450254	3	64200005	*PRICKLE2*	0.012	0.0013	2.3×10^{-21}	1.3×10^{-17}
cg06753787	2	2.2E+08	*ZFAND2B*	0.0063	0.0007	3.2×10^{-21}	1.8×10^{-17}
cg18158306	12	1.33E+08	*FBRSL1*	0.0102	0.0011	6.2×10^{-21}	3.2×10^{-17}
cg19093370	17	17110180	*PLD6*	0.0198	0.0021	8.7×10^{-21}	4.4×10^{-17}
cg09182189	1	1709203	*NADK*	−0.0104	0.0011	2.0×10^{-20}	9.2×10^{-17}
cg18369990	2	1.13E+08	*FBLN7*	0.0116	0.0013	2.3×10^{-20}	1.1×10^{-16}
cg24578857	17	17110207	*PLD6*	0.02	0.0022	3.1×10^{-20}	1.4×10^{-16}
cg20408402	10	72362452	*PRF1*	0.0085	0.0009	7.6×10^{-20}	3.1×10^{-16}
cg04673446	22	39879951	*MGAT3*	0.006	0.0007	2.0×10^{-19}	8.0×10^{-16}
cg06803614	1	40133581	*NT5C1A*	−0.0088	0.001	2.1×10^{-19}	8.3×10^{-16}
cg16274678	1	1.54E+08	*TPM3; NUP210L*	−0.0152	0.0017	2.9×10^{-19}	1.1×10^{-15}
cg07286341	5	1.77E+08	*PDLIM7*	−0.0077	0.0009	3.4×10^{-19}	1.3×10^{-15}
cg20674424	3	1.87E+08	*MIR1248; EIF4A2; SNORA81*	−0.0091	0.001	4.2×10^{-19}	1.5×10^{-15}
cg02279625	15	78384520	*SH2D7*	0.0105	0.0012	4.8×10^{-19}	1.7×10^{-15}
cg03485667	16	75143200	*ZNRF1*	−0.0168	0.0019	5.0×10^{-19}	1.8×10^{-15}
cg03531211	6	32920102	*HLA-DMA*	−0.0108	0.0012	7.5×10^{-19}	2.5×10^{-15}
cg09940677	14	1.03E+08	*CDC42BPB*	0.0081	0.0009	1.0×10^{-18}	3.2×10^{-15}

*CpG sites without gene names are intergenic. These are all included in all the analyses; †Coef stands for regression coefficients; ‡Not previously discovered by other studies

Table 3: Twenty-five most statistically significant CpG sites that were associated with former vs. never smoker status

Probe ID	Chr	Location	Gene Symbol*	Coef†	S.E.	P	FDR
cg01940273	2	2.33E+08	-	−0.0234	0.0013	9.6 × 10−73	1.8 × 10−68
cg25189904	1	68299493	*GNG12*	−0.0283	0.0021	3.5 × 10−40	3.3 × 10−36
cg12803068	7	45002919	*MYO1G*	0.0191	0.0017	9.3 × 10−31	5.8 × 10−27
cg19572487	17	38476024	*RARA*	−0.0159	0.0014	2.2 × 10−30	1.0 × 10−26
cg11554391	5	321320	*AHRR*	−0.0091	0.0008	1.0 × 10−28	3.9 × 10−25
cg05951221	2	2.33E+08	-	−0.0396	0.0036	1.1 × 10−27	3.2 × 10−24
cg23771366	11	86510998	*PRSS23*	−0.0167	0.0015	1.2 × 10−27	3.2 × 10−24
cg26764244	1	68299511	*GNG12*	−0.0119	0.0011	2.3 × 10−27	5.4 × 10−24
cg05575921	5	373378	*AHRR*	−0.0406	0.0038	8.2 × 10−27	1.7 × 10−23
cg11660018	11	86510915	*PRSS23*	−0.0157	0.0015	4.3 × 10−26	8.1 × 10−23
cg21566642	2	2.33E+08	-	−0.0434	0.0041	1.0 × 10−25	1.7 × 10−22
cg11902777	5	368843	*AHRR*	−0.0063	0.0006	2.8 × 10−25	4.3 × 10−22
cg26850624	5	429559	*AHRR*	0.0118	0.0011	3.1 × 10−25	4.4 × 10−22
cg03636183	19	17000585	*F2RL3*	−0.0267	0.0026	8.9 × 10−25	1.2 × 10−21
cg15693572	3	22412385	-	0.019	0.0019	1.5 × 10−23	1.9 × 10−20
cg17924476	5	323794	*AHRR*	0.0148	0.0016	4.0 × 10−20	4.7 × 10−17
cg12513616	5	1.77E+08	-	−0.0072	0.0008	2.4 × 10−19	2.7 × 10−16
cg07339236	20	50312490	*ATP9A*	−0.0062	0.0007	1.4 × 10−18	1.4 × 10−15
cg06126421	6	30720080	-	−0.0365	0.0042	3.0 × 10−18	3.0 × 10−15
cg14624207	11	68142198	*LRP5*	−0.0070	0.0008	5.0 × 10−18	4.7 × 10−15
cg00706683	2	2.33E+08	*ECEL1P2*	0.0101	0.0012	1.4 × 10−17	1.2 × 10−14
cg23351584	11	86512100	*PRSS23*	−0.0048	0.0006	7.0 × 10−17	6.0 × 10−14
cg02583484	12	54677008	*HNRNPA1*	−0.0062	0.0008	1.0 × 10−15	8.5 × 10−13
cg05302489	6	31760426	*VARS*	0.0079	0.001	2.5 × 10−15	2.0 × 10−12
cg01442064	4	5713450	*EVC*	−0.0055	0.0007	3.3 × 10−15	2.4 × 10−12

*CpG sites without gene names are intergenic. These are all included in all the analyses, †Coef stands for regression coefficients

Table 4: The top 36 most statistically significant CpG sites that did not return to never-smoker levels 30 years after smoking cessation in the FHS (N=2,648)

Probe ID	Chr	Location	Gene Symbol	P
cg05951221	2	2.33E+08	-	3.2 × 10−15
cg06644428	2	2.33E+08	-	1.2 × 10−14
cg05575921	5	373378	*AHRR*	6.5 × 10−14
cg21566642	2	2.33E+08	-	8.6 × 10−10
cg03636183	19	17000585	*F2RL3*	5.7 × 10−7
cg06126421	6	30720080	-	1.3 × 10−6
cg01940273	2	2.33E+08	-	1.9 × 10−6
cg23771366	11	86510998	*PRSS23*	3.1 × 10−6
cg17272563	6	32116548	*PRRT1*	4.4 × 10−6
cg23916896	5	368804	*AHRR*	1.3 × 10−5
cg11660018	11	86510915	*PRSS23*	1.3 × 10−5
cg08118908	16	15787920	*NDE1*	3.0 × 10−5
cg13937905	12	53612551	*RARG*	1.5 × 10−4
cg24172324	2	2.32E+08	-	1.7 × 10−4
cg10780313	6	33501379	-	2.0 × 10−4
cg14027333	6	32116317	*PRRT1*	2.1 × 10−4
cg11245297	19	8117898	*CCL25*	2.1 × 10−4
cg01692968	9	1.08E+08	-	3.1 × 10−4
cg00706683	2	2.33E+08	*ECEL1P2*	3.4 × 10−4
cg25317941	2	2.33E+08	*ECEL1*	4.0 × 10−4
cg25189904	1	68299493	*GNG12*	4.0 × 10−4
cg14179389	1	92947961	*GFI1*	4.7 × 10−4
cg13641317	3	1.27E+08	-	4.9 × 10−4
cg19847577	15	29213748	*APBA2*	5.1 × 10−4
cg14239618	7	1.1E+08	-	5.8 × 10−4
cg25955180	6	32116538	*PRRT1*	6.3 × 10−4

Probe ID	Chr	Location	Gene Symbol	P
cg00774149	3	52255721	*TLR9*	6.4 × 10−4
cg21351392	6	1.62E+08	*AGPAT4*	7.1 × 10−4
cg11902777	5	368843	*AHRR*	7.6 × 10−4
cg07251887	17	73641809	*LOC100130933; RECQL5*	7.7 × 10−4
cg19382157	7	2124566	*MAD1L1*	8.9 × 10−4
cg19925780	1	1.02E+08	-	1.1 × 10−3
cg03679544	6	1.56E+08	*TIAM2*	1.1 × 10−3
cg08559712	20	16030674	*MACROD2*	1.3 × 10−3
cg09837977	7	1.11E+08	*LRRN3; IMMP2L*	1.3 × 10−3
cg00931843	6	1.55E+08	*TIAM2*	1.4 × 10−3

*CpG sites without gene names are intergenic. These are all included in all the analyses.

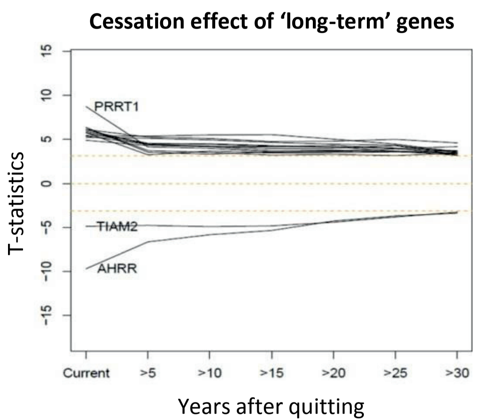


Figure 1. Trajectories of CpG sites that did not return to never-smoker levels within 30 years after cessation.

Enrichment analysis for genes identified in GWAS of smoking related phenotypes
To identify potential relevance of the differentially methylated genes to smoking-related phenotypes, we determined whether these genes had been associated with smoking-related phenotypes in the NHGRI-EBI GWAS Catalog[28] (accessed November 2, 2015). The catalog contained 9,777 SNPs annotated to 7,075 genes associated with 865 phenotypes at $p \leq 5x10^{-8}$. Of the 7,201 genes (mapped by 18,760 CpG sites) significantly differentially methylated in current versus never smokers, we found overlap with 1,791 genes (4,187 CpGs are mapped to these) associated in GWAS with 700 phenotypes (enrichment $p=2.4 \times 10^{-52}$). We identified smoking-related traits using the 2014 US Surgeon General's (USSG) report[1]. Enrichment results for a selection of smoking-related phenotypes including coronary heart disease (CHD) and its risk factors, various cancers, inflammatory diseases, osteoporosis, and pulmonary traits, are available in Table 5. We also performed the same enrichment analysis on the 2,568 CpGs associated with former versus never smoking status. We identified enrichment for CHD, pulmonary traits, and some cancers (Table 5). More detailed results are available in Supplementary Tables 8 and 9. Differentially methylated genes in relation to smoking status that are associated in GWAS with CHD or CHD risk factors are available in Supplementary Table 10. We also performed enrichment analyses on phenotypes that have no clear relationships to smoking, such as male pattern baldness (p=0.0888), myopia (p=0.1070), thyroid cancer (p=0.2406), and testicular germ cell tumor (p=0.3602) and did not find significant enrichment.

Enrichment analysis for genomic features
We examined the differentially methylated CpGs with respect to localization to different genomic regions including CpG islands, gene bodies, known differentially methylated regions, and sites identified as likely to be functionally important in the ENCODE project such as DNAse1 hypersensitivity sites and enhancers (refer to the Methods section for details). We performed this analysis separately for the CpGs related to current smoking and past smoking (Supplementary Table 11). Trends were similar for the two sets of CpGs, although the power to identify enrichment was much greater for the larger set of 18,760 CpGs related to current smoking. There was no enrichment for CpG islands. In contrast, significant enrichment was observed for island shores, gene bodies, DNAse1 hypersensitivity sites, and enhancers.

Table 5. Enrichment of CpGs for genome-wide association study (GWAS) phenotypes that are regarded as causally related to cigarette smoking

GWAS Phenotype	Enrichment p-value
Current vs. never smoking	
Coronary heart disease (CHD) and Stroke	0.0028
Ischemic stroke	0.0095
CHD risk factors	1.2×10^{-12}
Blood pressure / hypertension	8.1×10^{-6}
Diastolic blood pressure	6.1×10^{-5}
Systolic blood pressure	0.0008
Hypertension	0.015
Lipids	2.9×10^{-5}
High density lipoprotein (HDL)	0.0009
Type 2 diabetes	0.0106
Rheumatoid arthritis (RA)	2.9×10^{-5}
Bone mineral density (BMD) and osteoporosis	0.0467
All pulmonary traits	2.8×10^{-6}
All chronic obstructive pulmonary disease (COPD)	0.0295
Moderate to severe COPD	0.0156
Pulmonary function	0.0044
Crohn’s Disease	9.5×10^{-7}
Primary biliary cirrhosis	3.4×10^{-6}
Inflammation bowel disease	3.5×10^{-5}
Ulcerative colitis	9.8×10^{-5}
All cancer	8.0×10^{-15}
Lung adenocarcinoma	0.0015
Colorectal cancer	0.0014
Former vs. never smoking	
CHD risk factors	7.6×10^{-5}
Blood pressure / hypertension	5.8×10^{-5}
Diastolic blood pressure	0.0021
Systolic blood pressure	0.0002
Hypertension	0.0023
Rheumatoid arthritis (RA)	6.3×10^{-5}
All pulmonary traits	0.0217
Inflammation bowel disease	5.2×10^{-6}
Crohn’s Disease	0.0064
All cancer	7.8×10^{-6}

Transcriptomic integration

Of the 18,760 statistically significant CpG sites associated with current smoking in the meta-analysis, 1,430 were significantly associated in *cis* with the expression of 924 genes at FDR<0.05 (enrichment p=3.6 x 10^{-215}, Supplementary Table 12) using whole blood samples from 2,262 Framingham Heart Study participants. Of these, 424 CpGs associated with the expression of 285 genes were replicated at FDR<0.0001 in 1,264 CD14+ samples from the Multi-Ethnic Study of Atherosclerosis (MESA)[35]. These genes are associated with pathways similar to those described earlier (Supplementary Table 13).

Comparison between African ancestry and European ancestry

Meta-analysis of the current versus never smokers in 11 cohorts with participants of European ancestry (N=6,750 subjects) yielded 10,977 CpGs annotated to 4,940 genes at FDR<0.05. Meta-analysis the results of the smaller dataset of four cohorts with African ancestry participants (N=2,639) yielded 3,945 CpGs annotated to 2,088 genes at FDR<0.05. The effect estimates of the CpGs significant in at least one ancestry (12,927 CpGs) were highly correlated in the combined group of individuals of either ancestry (Spearman ρ=0.89). The results by ancestry are shown in Supplementary Table 14.

We performed the same ancestry-stratified analysis on former versus never smokers (Supplementary Table 15). Meta-analysis of the results of European ancestry participants yielded 2,045 CpG sites annotated to 1,081 genes at FDR<0.05. Meta-analysis of the results of African ancestry participants yielded 329 CpG sites annotated to 178 genes at FDR<0.05. The effect estimates of the union of CpGs significant in at least one ancestry (2,234 CpGs) were correlated in the combined group of individuals of either ancestry (Spearman ρ=0.75). Of note, one of CpG sites showing differential methylation in ancestry, cg00706683, mapped to gene *ECEL1P2*, did not return to never-smoker levels 30 years after smoking cessation (Table 4).

To more directly compare results by ethnicity removing the effect of better statistical power in the larger European ancestry sample size, we performed a meta-analysis on subset of European ancestry cohorts: the Framingham Heart Study, Rotterdam Study, and KORA, such that the total number of smokers, the major determinant of power, would match that of African ancestry cohorts. In this subset, similar correlations of the effect estimates were observed as in the complete analyses suggesting that the differences in number of statistically significant CpGs are indeed due to better power in the European ancestry cohorts (Spearman ρ=0.87 and 0.79 for current versus never smokers and former versus never smokers, respectively).

Cell type adjustment

We adjusted our main analyses for white blood cell fractions, in studies based on either whole blood or leukocytes from the buffy coat of whole blood, either measured or

using a published method[20]. Reassuringly, results before and after cell type adjustment were highly comparable. The correlation of regression coefficients before and after adjustment is 0.85 for the current *vs*. never smoker analysis (Supplementary Figure 7). Similarly for the analysis of former versus never smokers the effect estimates were highly correlated before and after adjustment (ρ=0.93; Supplementary Figure 8). In addition, in two cohorts we had results from specific cell fractions - CD4+ cells in GOLDN and CD14+ cells in MESA. The correlation of results between buffy coat and CD4+ or CD14+ for former versus never smokers are generally high ($\rho > 0.74$; Supplementary Table 16).

Methylation profile across CpG sites

We assessed methylation profile in FHS cohort as a representative cohort in the study. The profile of all 485,381 analyzed CpG sites can be found in Supplementary Figure 9. The profile across 18,760 CpG sites significantly associated with current *vs*. never smoking status can be found in Supplementary Figure 10. These plots indicate that most CpG sites with less dynamic range are largely not statistically significant in our results.

DISCUSSION

We performed a genome-wide meta-analysis analysis of blood-derived DNA methylation in 15,907 individuals across 16 cohorts and identified broad epigenome-wide impact of cigarette smoking, with 18,760 statistically significant CpGs (FDR<0.05) annotated to over 7,000 genes, or roughly a third of known human genes. These genes in turn affect multiple molecular mechanisms and are implicated in smoking-related phenotypes and diseases. In addition to confirming previous findings from smaller studies, we detected over 16,000 novel differentially methylated CpGs in response to cigarette smoking. Many of these genes have not been previously implicated in the biologic effects of tobacco exposure. The large number of genes implicated in this well powered meta-analysis might on first glance raise concerns about false positives. However, on further consideration, given the widespread impact of smoking on disease outcomes across many organ systems and across the lifespan[1], the identification of a large number of genes at genome wide significance is not surprising. In addition, our findings are robust and consistent across all 16 cohorts (Supplementary Tables 2 and 5) because we accounted for inter-study variability by using random effect meta-analyses, which is conservative when heterogeneity is present[36]. The implicated genes are mainly involved in molecular machineries, such as transcription and translation. Furthermore, differential methylation of a subset of CpGs persisted, often for decades, following smoking cessation.

We found that genes differentially methylated in relation to smoking are enriched for variants associated in GWAS with smoking–related diseases[1] including,

osteoporosis, colorectal cancers, chronic obstructive pulmonary disease, pulmonary function, cardiovascular disease (CVD) and rheumatoid arthritis. We find it noteworthy that there is enrichment of smoking-associated CpGs for genes associated with rheumatoid arthritis because DNA methylation is one of the proposed molecular mechanisms underlying this disease [37]. It is also interesting that the most significant association of smoking with methylation was for the gene *HIVEP3* (a.k.a. Schnurri3), the mammalian homolog of the Drosophila zinc finger adapter protein Shn[38]. This gene regulates bone formation, an important determinant to osteoporosis, which was one of the enriched GWAS phenotypes.

When we examined time since smoking cessation, we found that the majority of the differentially methylated CpG sites observed in analysis of current versus never smokers returned to the level of never-smokers within five years of smoking cessation. This is consistent with the fact that risks of many smoking-related diseases revert to nonsmoking levels within this period of time. Our results also indicate that cigarette smoking induces long-lasting alterations in DNA methylation at some CpGs. While speculative, it is possible that persistent methylation changes at some loci might contribute to risks of some conditions that remain elevated after smoking cessation.

In all but two of our 14 cohorts DNA was extracted from the entire circulating leukocyte population. Thus there is the possibility of confounding by the effects of smoking on differential cell counts. We attempted to adjust for cell type and found that results were generally little changed by the adjustment.

Our significant results are highly enriched for CpG sites associated with the expression of nearby genes (*i.e.*, in *cis*) even though a single measurement of gene expression in blood is probably subject to considerably more within-subject variability than DNA methylation,[39] limiting our ability to find correlations. Differential DNA methylation at many of the CpGs we identified in relation to smoking status may have a functional impact on nearby gene expression. Our analysis of genomic regions further supports the potential functional impact of our findings on gene expression. We demonstrated enrichment for sites with greater functional impact such as island shores, gene bodies, DNAse1 hypersensitivity sites, and enhancers, whereas we found no enrichment for CpG islands. These results reinforce previous findings showing that island shores, enhancers, and DHS sites are more dynamic (*i.e.*, susceptible to methylation changes) than CpG islands[40], which may be more resistant to abrupt changes in DNA methylation in response to environmental exposures[41]. Thus our results suggest that many of the smoking-associated CpG sites may have regulatory effects.

While identification of changes in methylation patterns may suggest mechanisms by which exposure to tobacco smoke exerts its effects on several disease processes, DNA methylation profiles can also serve as biomarkers of exposure to tobacco smoke. Cotinine is a biomarker only of recent smoking; DNA methylation

signals have the potential to serve as robust biomarkers of past smoking history[9,42]. Indeed, several studies have identified several of such markers[5,42,43]. The large number of persistently modified CpGs may be useful to develop even more robust biomarkers to objectively quantify long-term cigarette smoking exposure for prediction of risk for health outcomes in settings where smoking history is not available or is incomplete as well as to validate self-reported never smoker status. Further, our analyses of both former and current smokers show dose-dependent effects at a number of CpGs (Supplementary Tables 3 and 7). Methylation based biomarkers could be informative for investigating dose response relationships with disease endpoints. This is useful because smokers often underreport the amount of smoking, both current and historical.

It is possible that smoking related conditions or correlated exposures may contribute to some of the methylation signatures identified. However, our studies are nearly all population based studies composed of predominantly healthy individuals, not selected for smoking related disease. Given the number, strength and robustness to replication of findings for smoking across the literature and among our diverse cohorts from various countries the likelihood that these are confounded by other exposures or conditions related to smoking is greatly reduced.

There several potential limitations to our study. First, the cross-sectional design limits our ability to study the time course of smoking effects. In addition, we analyzed methylation in DNA samples from blood, which is readily accessible. Although we demonstrated that blood derived DNA reveals a strong and robust signature of cigarette smoking exposure, studies in target tissues for smoking-related diseases (*e.g.*, heart and lung) would be of additional interest. In addition, our analyses could not distinguish smoking's direct effects from its indirect effects due to smoking-induced changes in cell metabolism, organ function, inflammation, or injury that could in turn influence methylation. However, this is the largest examination to date of the effects of smoking on DNA methylation with 16 studies from different countries contributing.

In conclusion we identify an order of magnitude more sites differentially methylated in relation to smoking across the genome than have been previously seen. Many of these signals persist long after smoking cessation providing potential biomarkers of past smoking history. These findings may provide new insights into molecular mechanisms underlying the protean effects of smoking on human health and disease.

Supplementary Table 1. Detailed participant characteristics by cohort. [Additional Information accompanies this paper at *http://circgenetics.ahajournals.org/content*.]

					N			Males (%)			Age (Mean/SD)			BMI (Mean/SD)				Pack Years (Mean/SD)	
Cohort	Ethnicity	Sample Origin	CpG at FDR<0.05 Current vs. Never	CpG at FDR<0.05 Former vs. Never	Current	Former	Never	Current	Former	Never	Current	Former	Never	Current	Former	Never	Cell types	Current	Former
ARIC	AA	BC	7,794	61	720	866	1262	49.7	48.3	20.7	56.2 5.8	57.2 5.9	56.5 5.8	27.9 5.8	30.2 6.2	31.3 6.3	H	25.46 21.97	16.10 20.94
GTP	AA	WB	81	0	94	64	128	37.2	26.6	24.2	43.4 11.7	44.6 13.1	38.4 12.9	33.0 8.3	33.3 6.8	32.9 8.7	H	NA	NA
CHS AA	AA	WB	20	0	30	66	96	46.7	53.0	18.8	70.4 4.9	72.7 5.2	74.1 5.6	25.7 3.7	28.7 4.5	29.7 5	H	28.5 18.3*	22.5 19.8*
GENOA	AA	BC	127	6	77	111	232	50.6	42.3	15.1	58.7 7.9	61.7 7.7	61.0 7.4	28.2 5.6	31.0 6.1	31.4 6.0	H	29.5 17.4	23.1 22.9
FHS	EA	WB	14,318	199	274	1538	836	51.8	51.1	33.7	62.5 7.8	67.4 8.5	65.8 9.6	27.7 5.6	28.5 5.2	27.9 5.4	H	37.8 23.5	7.3 12.3
KORA F4	EA	WB	6,513	42	262	782	753	54.2	60.5	34.5	57.0 7.0	61.1 9.0	62.1 8.9	27.1 4.8	28.6 5	28 4.6	H	33.39 23.95	21.59 24.33
GOLDN	EA	CD4+	51	10	73	216	703	49.3	57.9	44.5	44 13	56 14	47 17	28.1 5.8	28.9 5.0	28.1 5.8	NA	18.57 17.86	18.09 22.44
LBC 1921	EA	WB	73	9	31	213	201	35.5	46.0	33.3	79.2 0.5	79.1 0.6	79.1 0.6	24.9 4.0	26.4 3.7	26.1 4.4	H		
LBC 1936	EA	WB	476	24	103	385	432	45.6	58.4	44.7	69.5 0.7	69.6 0.8	69.5 0.8	25.9 4.5	28.5 4.5	27.6 4.1	H		
NAS	EA	WB	176	15	26	431	187	100.0	100.0	100.0	68.2 6.1	72.3 6.6	73.8 7.3	28.2 4.7	28.3 4.1	27.4 3.8	H	59.8 28.6	27.7 24

RS	EA	WB	786	7	169	307	210	45.0	47.2	37.1	58.0 6.8	61.6 8.6	59.2 8.4	26.3 3.9	28.0 4.5	27.7 4.9	M	30.8 22.2	19.6 20.9
Inchianti	EA	WB	613	7	50	170	288	52.0	72.4	27.8	58.9 16.8	74.3 12.2	72.6 16.6	26.2 4.6	27.5 3.8	26.8 4.4	M	28 20.9	22.8 19.2
CHS EA	EA	WB	2	3	23	76	85	34.8	57.9	34.1	74.1 4.2	75.6 5.1	76.8 5.2	26.8 5.2	27 4.9	27 4.8	H	NA	NA
EPIC-Norfolk	EA	BC	844	11	191	463	529	53.9	65.0	34.6	58.3 8.4	61.4 8.7	59.9 8.9	25.9 4.2	27.7 4.4	27.1 4.4	H		
MESA	EA+AA+HIS	CD14+	3,243	17	114	640	502	55.3	55.8	37.8	65 8	70 9	70 10	28 6	30 5	30 6	NA	36 27	16 20
EPIC	EA	BC	0	0	196	190	512	0.0	0.0	0.0	48.9 8.8	51.3 8.9	54.0 8.6	25.0 4.8	25.5 4.7	26.2 4.3	H	NA	NA
Total					2433	6518	6956												

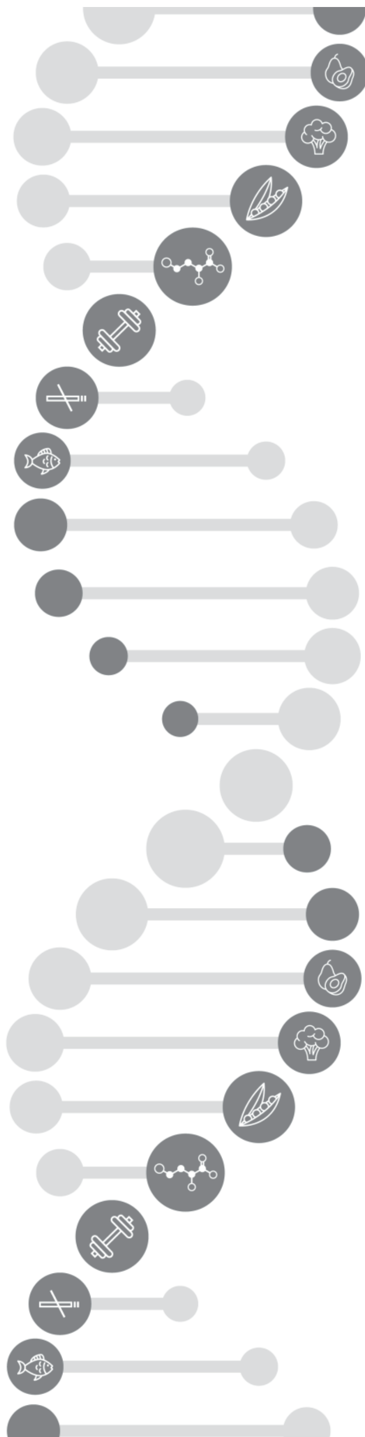
* For which pack years data are available; WB = Whole Blood; BC = Buffy Coats; H = Houseman, M = Measured

REFERENCES

1. National Center for Chronic Disease Prevention and Health Promotion (US) Office on Smoking and Health. The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General. Atlanta (GA): Centers for Disease Control and Prevention (US); 2014
2. World Health Organization. WHO global report on trends in prevalence of tobacco smoking. 2015.
3. Szarc vel Szic K, Declerck K, Vidaković M, Vanden Berghe W. From inflammaging to healthy aging by dietary lifestyle choices: is epigenetics the key to personalized nutrition? *Clin Epigenetics*. 2015;7:33.
4. Breitling LP, Yang R, Korn B, Burwinkel B, Brenner H. Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. *Am J Hum Genet*. 2011;88:450–457.
5. Breitling LP, Salzmann K, Rothenbacher D, Burwinkel B, Brenner H. Smoking, F2RL3 methylation, and prognosis in stable coronary heart disease. *European heart journal*. 2012;33:2841-8.
6. Wan ES, Qiu W, Baccarelli A, Carey VJ, Bacherman H, Rennard SI, et al. Cigarette smoking behaviors and time since quitting are associated with differential DNA methylation across the human genome. *Human molecular genetics*. 2012;21:3073-3082
7. Wan ES, Qiu W, Carey VJ, Morrow J, Bacherman H, Foreman MG, et al. Smoking Associated Site Specific Differential Methylation in Buccal Mucosa in the COPDGene Study. *Am J Respir Cell Mol Biol*. 2014;53:246-54.
8. Zeilinger S, Kühnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, et al. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. *PLoS ONE*. 2013;8:e63812.
9. Shenker NS, Ueland PM, Polidoro S, van Veldhoven K, Ricceri F, Brown R, et al. DNA Methylation as a Long-term Biomarker of Exposure to Tobacco Smoke: *Epidemiology*. 2013;24:712–716.
10. Shenker NS, Polidoro S, van Veldhoven K, Sacerdote C, Ricceri F, Birrell MA, et al. Epigenome-wide association study in the European Prospective Investigation into Cancer and Nutrition (EPIC-Turin) identifies novel genetic loci associated with smoking. *Hum Mol Genet*. 2013;22:843–851.
11. Guida F, Sandanger TM, Castagné R, Campanella G, Polidoro S, Palli D, et al. Dynamics of Smoking-Induced Genome-Wide Methylation Changes with Time Since Smoking Cessation. *Hum Mol Genet*. 2015;24:2349-59.
12. Qiu W, Wan E, Morrow J, Cho MH, Crapo JD, Silverman EK, et al. The impact of genetic variation and cigarette smoke on DNA methylation in current and former smokers from the COPDGene study. *Epigenetics*. 2015;10:1064–1073.
13. Gao X, Jia M, Zhang Y, Breitling LP, Brenner H. DNA methylation changes of whole blood cells in response to active smoking exposure in adults: a systematic review of DNA methylation studies. *Clin Epigenetics*. 2015;7:113.
14. Shah S, Bonder MJ, Marioni RE, Zhu Z, McRae AF, Zhernakova A, et al. Improving Phenotypic Prediction by Combining Genetic and Epigenetic Associations. *Am J Hum Genet*. 2015;97:75–85.
15. Beane J, Sebastiani P, Liu G, Brody JS, Lenburg ME, Spira A. Reversible and permanent effects of tobacco smoke exposure on airway epithelial gene expression. *Genome Biology*. 2007;8:R201.
16. Wauters E, Janssens W, Vansteenkiste J, Decaluwé H, Heulens N, Thienpont B, et al. DNA methylation profiling of non-small cell lung cancer reveals a COPD-driven immune-related signature. *Thorax*. 2015;70:1113-22.

17. Garg AX, Hackam D, Tonelli M. Systematic Review and Meta-analysis: When One Study Is Just not Enough. *Clinical Journal of the American Society of Nephrology*. 2008;3:253–260.
18. Pidsley R, Y Wong CC, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics*. 2013;14:293.
19. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics*. 2013;29:189–196.
20. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*. 2012;13:86.
21. Dick KJ, Nelson CP, Tsaprouni L, Sandling JK, Aïssi D, Wahl S, et al. DNA methylation and body-mass index: a genome-wide analysis. *Lancet*. 2014;383:1990–1998.
22. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*. 2015;67:1-48.
23. R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: 2010.
24. Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software*. 2010;36:1-48.
25. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *JSSRB*. 1995;57:289–300.
26. Joubert BR, Felix JF, Yousefi P, Bakulski KM, Just AC, Breton C, et al. DNA Methylation in Newborns and Maternal Smoking in Pregnancy: Genome-wide Consortium Meta-analysis. *Am J Hum Genet*. 2016;98:680–696.
27. Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, et al. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci USA*. 2004;101:6062–6067.
28. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci USA*. 2009;106:9362–9367.
29. Vazquez AI, Bates DM, Rosa GJM, Gianola D, Weigel KA. Technical note: an R package for fitting generalized linear mixed models in animal breeding. *J Anim Sci*. 2010;88:497–504.
30. Devlin B, Roeder K. Genomic control for association studies. *Biometrics*. 1999;55:997–1004.
31. Martey CA, Baglole CJ, Gasiewicz TA, Sime PJ, Phipps RP. The aryl hydrocarbon receptor is a regulator of cigarette smoke induction of the cyclooxygenase and prostaglandin pathways in human lung fibroblasts. *Am J Physiol Lung Cell Mol Physiol*. 2005;289:L391-399.
32. Teschendorff AE, Yang Z, Wong A, Pipinikas CP, Jiao Y, Jones A, et al. Correlation of Smoking-Associated DNA Methylation Changes in Buccal Cells With DNA Methylation Changes in Epithelial Cancer. *JAMA Oncology*. 2015;1:476-485.
33. Campesi I, Carru C, Zinellu A, Occhioni S, Sanna M, Palermo M, et al. Regular cigarette smoking influences the transsulfuration pathway, endothelial function, and inflammation biomarkers in a sex-gender specific manner in healthy young humans. *Am J Transl Res*. 2013;5:497–509.
34. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. 2005;102:15545–15550.
35. Liu Y, Ding J, Reynolds LM, Lohman K, Register TC, De La Fuente A, et al. Methylomics of gene expression in human monocytes. *Hum Mol Genet*. 2013;22:5065–5074.

36. Han B, Eskin E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am J Hum Genet*. 2011;88:586–598.
37. Liu Y, Aryee MJ, Padyukov L, Fallin MD, Hesselberg E, Runarsson A, et al. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nat Biotechnol*. 2013;31:142–147.
38. Jones DC, Wein MN, Oukka M, Hofstaetter JG, Glimcher MJ, Glimcher LH. Regulation of adult bone mass by the zinc finger adapter protein Schnurri-3. *Science*. 2006;312:1223–1227.
39. Suderman M, Pappas JP, Borghol N, Buxton JL, McArdle WL, Ring SM, et al. Lymphoblastoid cell lines reveal associations of adult DNA methylation with childhood and current adversity that are distinct from whole blood associations. *Int J Epidemiol*. 2015;44:1331-40.
40. Ziller MJ, Gu H, Müller F, Donaghey J, Tsai LT-Y, Kohlbacher O, et al. Charting a dynamic DNA methylation landscape of the human genome. *Nature*. 2013;500:477–481.
41. Ivanova E, Chen J-H, Segonds-Pichon A, Ozanne SE, Kelsey G. DNA methylation at differentially methylated regions of imprinted genes is resistant to developmental programming by maternal nutrition. *Epigenetics*. 2012;7:1200–1210.
42. Zhang Y, Schöttker B, Florath I, Stock C, Butterbach K, Holleczek B, et al. Smoking-Associated DNA Methylation Biomarkers and Their Predictive Value for All-Cause and Cardiovascular Mortality. *Environ Health Perspect*. 2015; 124:67-74.
43. Zhang Y, Yang R, Burwinkel B, Breitling LP, Brenner H. F2RL3 Methylation as a Biomarker of Current and Lifetime Smoking Exposures. *Environmental Health Perspectives*. 2013;122:131-7.



PART D

Discussion and summary

CHAPTER 9: GENERAL DISCUSSION

The aim of this thesis was to investigate the effects of key factors from the methyl-group metabolism on genome-wide DNA methylation in a site-specific manner. In addition, we also aimed to investigate the effects of other lifestyle factors such as nutrition and cigarette smoking on DNA methylation. We conducted several epigenome-wide association studies (EWASs) in blood samples of a population-based cohort, the Rotterdam Study (RS) and collaborated with other cohorts with similar data in a consortium based effort, to increase power and robustness of our findings.

MAIN FINDINGS

Methyl-group Metabolism and B-vitamins

In Chapter 4, we describe that plasma Hcy was associated with subtle differences in DNA methylation of blood DNA, since we identified a small number of differentially methylated positions (3 DMPs) associated with Hcy in leukocytes of 2,035 individuals. Regional analysis identified 68 differentially methylated regions (DMRs) associated with Hcy. In Chapter 5, genetically elevated Hcy levels assessed by the *MTHFR* 677C>T variant and combined weighted genetic risk score (GRS) of 18 previously studied Hcy-associated variants, did not show widespread changes in DNA methylation; identifying a few DMPs in *cis* (< 1Mb upstream or downstream of the SNP) and one DMP in *trans* (> 1Mb upstream or downstream of the SNP) associated with the *MTHFR* 677C>T variant and five trans-CpGs associated with the GRS. In Chapter 6, we conducted an interaction study of Hcy and the *MTHFR* 677C>T variant to investigate their combined effect on genome-wide DNA methylation in leukocytes of 1280 individuals. We identified 13 DMPs associated with [Hcy x *MTHFR* 677C>T]. This included four cis-DMPs at the *AGTRAP-MTHFR-NPPA/B* gene cluster on chromosome 1, which were also described in Chapter 5, indicating that this association is more likely driven by MTHFR 677C>T variant since they are cis-DMPs for the MTHFR 677C>T variant. In Chapter 7, in leukocytes of 5,841 individuals, we identified six novel DMPs and 74 DMRs significantly associated with dietary folate intake, and 29 DMRs significantly associated with dietary vitamin B12 intake. No DMPs were found associated with dietary vitamin B12 intake.

In conclusion, we investigated several important markers of one-carbon metabolism and found only subtle differences in DNA methylation that were of modest statistical significance. When we examined overlap between the identified methylation markers with the various factors from the methyl-group metabolism, we observed that all identified CpGs were specific for the marker under study and no overlap was found.

Lifestyle factors

In Chapter 8, we focused on association between cigarette smoking as a lifestyle factor and DNA methylation in leukocytes assessed in 15,907 individuals. We identified

widespread changes in methylation: 18,760 DMPs when comparing current versus never smokers, and 2,623 DMPs when comparing former versus never smokers. This indicates a pattern of persistent altered methylation, with attenuation for up to five years, after smoking cessation.

CONSIDERATIONS OF OUR MAIN FINDINGS

Despite the fact that the methyl-group metabolism is the key metabolism delivering methyl-groups for DNA-methylation, we did not find widespread effects of Hcy, Hcy-associated genetic variants and dietary folate and vitamin B12 on DNA methylation across the genome. By comparison, when we studied the relation between cigarette smoking and DNA methylation, we identified widespread differences. There could be several explanations for the observed minimal effects of methyl-group metabolism on methylation pattern, that are listed below:

Mildly elevated Homocysteine levels

Hcy concentrations between 15-30 µmol/L are referred to as mild or moderate hyperhomocysteinemia (HHcy), whereas concentrations between 30-100 µmol/L and >100 µmol/L are referred to as intermediate and severe HHcy, respectively [1-3]. The subtle findings in Chapter 4 and Chapter 6 may be explained by the generally low median concentrations of Hcy in each cohort, with only about 10.2–32.3% of individuals in the mild or moderate range of HHcy above 15 µmol/l. In other words, the range of 3.4-35.4 µmol/l of Hcy might be too low, indicating a still balanced methyl-group metabolism. It might be that more extreme deregulation of the methyl-group metabolism, as present in intermediate or severe HHcy, as occurs in severe folate or vitamin B12 deficiency, diseased individuals or inborn errors of metabolism, will lead to deregulation of DNA-methylation.

Mildly elevated Hcy concentrations might not have resulted in altered s-adenosylmethionine (SAM) and s-adenosylhomocysteine (SAH) levels in study participants. The ratio of SAM:SAH is frequently used as an indicator of cellular methylation potential, where SAM is the substrate and SAH is the product of essential methyltransferase reactions. A few human studies have suggested that elevated Hcy with corresponding increase in SAH is more strongly associated with global DNA hypomethylation, as compared to corresponding alteration in SAM [4, 5]. SAM and SAH was measured only in few previous studies, where they reported that individuals with mild, moderate or intermediately elevated Hcy levels showed increased SAM or increased SAH, with some showing corresponding decrease in SAM:SAH ratio [5-8].

We measured Hcy concentrations in plasma. Since methyl-group metabolism effectively takes place intracellularly, a double-blind trial by Smith et. al. [9] showed that FA supplementation significantly lowered plasma Hcy, but not intracellular Hcy. FA supplementation also did not lower intracellular SAM and SAH, which is possibly the

reason why global DNA methylation as measured by liquid chromatography-tandem mass spectrometry remained unaffected [9, 10].

Dietary data

Dietary intake of folate and vitamin B12 that are major determinants of B-vitamin status, was associated with subtle differences in genome-wide methylation profiles. In older individuals, vitamin B12 uptake is usually impaired as a consequence of issues related to absorption [11], health disorders, medications or lifestyle factors [12]. Given the fact that the mean age in the majority of our examined cohorts was over 60 years, it is highly likely that dietary vitamin B12 intake might not accurately reflect intracellular status and therefore, could result in misclassification of vitamin B12 status [13, 14]. Dietary data is also prone to measurement errors, which can compromise our ability to detect statistically significant associations. Furthermore, there were differences in nutrient intake ranges across cohorts which could be related to the number and type of food items used in the FFQs, and the types of food composition reference tables used to calculate the nutrient data. Except for those studies for which covariates were not present, in the analysis, adjustments were done for age, sex, body mass index (BMI), differential white blood cell (WBC) counts, smoking status, physical activity, B-vitamin supplement intake, and alcohol (grams per day or drinks per week) and coffee (grams per day or servings per day) consumption. All these adjustments could have reduced the amount of variability left in nutrient intake, resulting in reduced power. Notably, the findings from previous birth cohort studies identified much more DNA methylation changes in neonates in relation to maternal folate intake or levels, as compared to our study in elderly [15, 16]. It could be that the effect of low B-vitamins intake is not seen in the individual itself but that the effects of low B vitamins are transmitted to the next generation to show an effect on DNA [17, 18]. Moreover, fetuses are very sensitive for disturbances in the methyl-group metabolism during development due to increased placental and fetal demand for folate, and therefore, any insufficiencies could have major implications [19-21]. Lastly, there could be a *threshold* effect for a possible association, i.e., there is a possibility that regulation of methylation is very tight and that mild differences in the metabolism can be compensated for, but not so for more extreme deregulation. This could mean that in our cohorts the folate and vitamin B12 status were consistently sufficiently optimal to prevent changes in DNA methylation patterns.

Cohort heterogeneity and sample size

Hcy concentrations are associated with different pathophysiologies and it should be noted that in our studies, the study population were either population based, or consisted of cases with venous thromboembolism, or have a mildly increased risk to develop Type 2 diabetes and/or cardiovascular disease. Individuals with Type 2

diabetes have a different metabolism due to insulin resistance and impaired insulin secretion [22]. DNA methylation of insulin related and unrelated genes have been previously associated with insulin resistance and Type 2 diabetes [22, 23]. In addition, SNPs associated with type 2 diabetes [24] and/or cardiovascular diseases [25, 26] might affect DNA methylation variability. Therefore, this clinical heterogeneity between cohorts could have hindered identifying DMPs that are associated with Hcy.

A large sample size in EWAS studies is essential to improve statistical power in order to detect subtle changes and can possibility compensate for cohort heterogeneity. We performed an EWAS in relation to Hcy in 2,035 individuals and in relation to folate and B12 in 1,280 individuals and were not able to show large difference. In addition, we studied the effect of *MTHFR* 677C>T and other Hcy-associated DNA variants on methylation in a sample size of 9,894 individuals, we again did not find large differences in methylation. This genetic study strongly suggest that there is no large difference in methylation patterns of circulating leukocytes in individuals with mildly elevated homocysteine levels.

RNA or protein methylation

Hcy is an important intermediate of the methyl-group metabolism that mediates the transfer of methyl groups to DNA as well as to RNA and proteins. Recent studies have shown that RNA methylation at the 6 methyl position of adenine (m^6A) and 5 methyl position of cytosine (5-mC) regulates processes such as gene expression, RNA splicing, RNA stability, RNA degradation, mRNA translation and development, the aberrant methylation of which showing its implications in health and diseases [27-30]. Furthermore, protein methylation particularly on the nitrogen side-chains of the arginine and lysine residues are also known to be involved in numerous cellular processes such as DNA damage repair, RNA metabolism and gene expression [31-33]. It could be that the effects of Hcy and other methyl-group metabolism factors are more specific on the methylation of RNA or proteins.

In an animal model, diet-induced HHcy altered global protein arginine methylation in a tissue-specific manner [34]. Plasma Hcy elevation is known to parallelly increase plasma SAH elevation leading to subsequent DNA hypomethylation [5]. When the effect of accumulation of intracellular SAH on both DNA methylation and protein arginine methylation was further investigated in cultured human umbilical vein endothelial cells, it was found that protein arginine methylation is more sensitive to SAH accumulation than DNA methylation, indicating that proteins are more prone to be hypomethylated than genomic DNA [35]. Large studies on the relation between (factors of the) methyl-group metabolism and methylation of RNA and/or proteins are currently missing, and it therefore remains to be discovered whether this relationship exists.

Tissue specificity

The most commonly studied tissues in molecular epidemiology are blood and buccal epithelial cells as they are easy to obtain and are readily available. We conducted EWASs of Hcy level, by *MTHFR* 677C>T genotype and for folate/vitamin B12 intake, in whole blood leukocytes. However these effects could be different across tissues such as heart, brain and liver, from which methylation profiles are difficult to obtain. Few studies have performed DNA methylome profiling of other human tissues [36-39]. Tissue specific methylation differences are also seen in genomic DNA and domain of the *H19* gene of liver, aorta, kidney and brain in mice with HHcy [40-42]. The tissue-specific methylation reflects tissue-specific regulation of RNA expression and the biological dynamics are explained by the turnover or maintenance of binding sites for important transcription factors [43]. Moreover, it has been shown that the methylation profile of blood is highly distinct and contains relatively low levels of methylation as compared to other somatic tissues such as buccal cells [44-48]. The difference in DNA methylation between blood as compared to buccal cells, was more prominent at regions of low CpG density than at medium or high CpG density, while buccal cells had more variable CpGs as compared to blood tissue [47]. This distinction of blood as compared to other tissues was suggested to be due to the fact that blood is not a solid tissue and that the components of blood are a part of the inflammatory and immune systems, which are functionally related to many diseases [46, 47, 49, 50]. The other readily available tissue, buccal cells showed a significantly larger overlap with hypomethylated regions in a large number of other tissues in one report [44]. Therefore, buccal cells might be better than blood for large-scale epidemiological studies. It is clear that using blood as a surrogate tissue necessitates careful interpretation of results and correlation with the tissue of interest [51]. Depending on the clinical interest, other tissues should be further investigated such as heart, liver, kidney and brain.

Cellular heterogeneity

DNA methylation measurements from whole blood leukocytes is a mix of different cell types with varying proportion of different white blood cell counts (WBCs). There are five types of WBCs which include neutrophils (~62%), eosinophils (~2.3%), basophil (~0.4%), lymphocytes (~30%) and monocytes (~5.3%). Neutrophils, eosinophils and basophils together form the granulocytes. Since DNA methylation can be specific for different cell types, its analysis might be confounded due to differences in interindividual composition of cell types [52-57]. Therefore, DNA methylation analysis requires cautious consideration of cell count adjustments in the model. In order to adjust for cellular heterogeneity, RS and most other studies used measured percentage of differential cell counts of granulocytes, lymphocytes and monocytes. Other studies used percentage of imputed cell counts derived from the houseman method [58], to

adjust for cellular heterogeneity. The Houseman method uses methylation profiles to statistically estimate relative proportions of cell type components in whole blood that includes granulocytes, monocytes and subtypes of lymphocytes that include the B cells, natural killer (NK) cells, and T cells. The R^2 correlation between measured and houseman imputed cell counts is 0.580, 0.645 and 0.259 in granulocytes, lymphocytes and monocytes, respectively, in a subset of 700 samples of the Rotterdam Study [Figure 1]. It is important to keep in mind that Houseman uses a reference-based algorithm that is based on only six reference individuals [52]. Therefore, its estimates are not fully optimal, and can introduce biases in association results.

Even after cell type adjustments, cellular heterogeneity could still be an issue due to several reasons. One, if a particular cell type is underrepresented, then a change in DNA methylation associated to it may not be detected, and can only be detected if DNA methylation in such cell type is measured separately [52, 59]. Two, in important cell types of blood, few CpGs in regulatory regions may have opposite methylation patterns between the cell types. Considering these interindividual differences between cells, cautious interpretation of the overall methylation results is needed [52]. Three, B cells show a separate cluster and the largest number of DMPs with its associated genes belonging to other pathways, as compared to other lymphocytes [52]. This distinct methylation profile of B cells should be considered, and adjusting for the percentage of total measured lymphocytes may not reduce this heterogeneity.

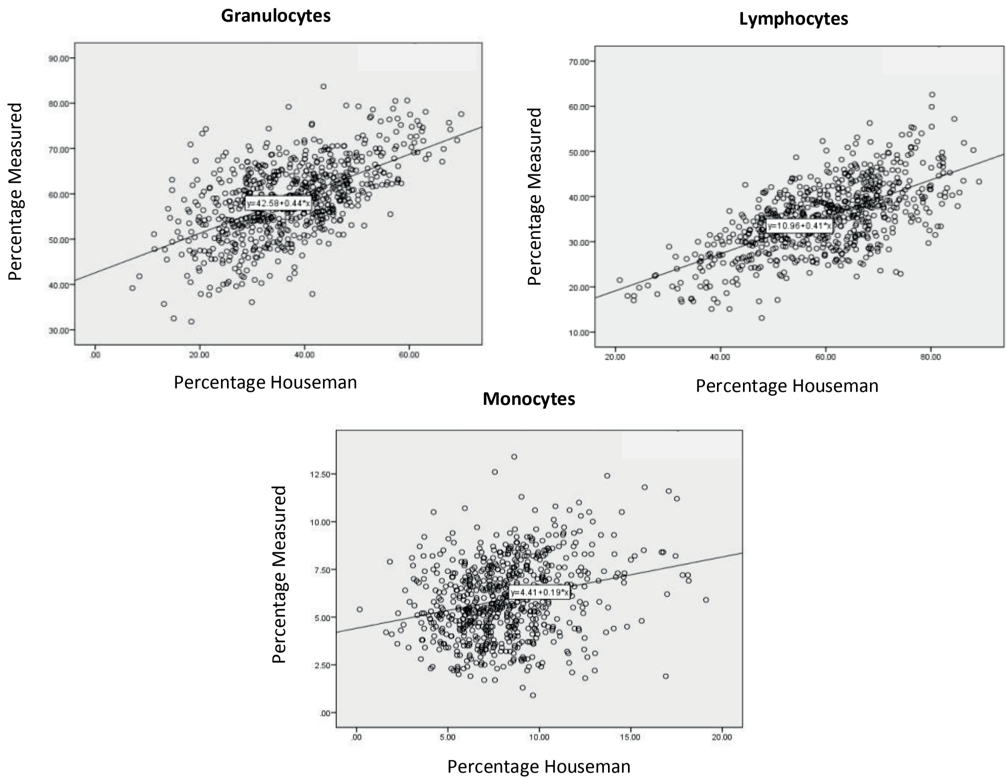


Figure 1. Correlation between measured and houseman imputed cell counts (N=700, blood samples of the Rotterdam Study). For granulocytes, lymphocytes and monocytes, the pearson correlation coefficient (R^2) was 0.580 (N=699, one outlier removed), 0.645 (N=700) and 0.259 (N=696, four outliers removed), respectively, at p <0.01.

Cigarette smoking

We observed a large epigenome-wide impact of smoking on DNA methylation pattern with a total of 18,760 significant DMPs annotated to over 7,000 genes. Why is the effect of smoking so much larger than Hcy and other methyl-group metabolism factors? One, Hcy and other methyl-group metabolism factors such as folate and vitamin B12 are associated with smoking [60-62] and with other lifestyle factors such as physical activity [63, 64], coffee consumption [65-67] and alcohol consumption [68, 69]. Therefore, Hcy, folate and vitamin B12 cannot be seen separately from lifestyle factors and that these factors might have masked some associations. Moreover, despite our inclusion of these known confounders as covariates in the B-vitamin EWAS models, the possibility of inter-correlations of these confounders with each other or other factors belonging to the causal pathway could nullify the findings. Two, in addition to the methyl-group metabolism, smoking could affect DNA methylation via few other known mechanisms [70] such as:

a) Cigarette smoke contains carcinogens such as arsenic, chromium, polycyclic aromatic hydrocarbons, formaldehyde, and nitrosamines that causes double-stranded breaks to the DNA causing DNA damage [71]. The enzyme DNMT1 is recruited to repair the damage in surviving cells which causes methylation of CpGs adjacent to the repaired nucleotides [71-73].

b) Cigarette smoke can alter DNA methylation through the effect of *nicotine* on DNMT1 expression. Nicotine binds to and activates the nicotinic acetylcholine receptors that increases intracellular calcium and activates the key transcription factor (cAMP response element-binding protein) of many genes with possible downregulation of DNMT1 [74-76].

c) Cigarette smoking increases expression of the DNA-binding protein Sp1 which protects CpG sites from being methylated [77-79].

d) Cigarette smoke contains carbon monoxide that competes with oxygen to bind to hemoglobin. This causes inadequate oxygenation of tissues in a condition called hypoxia [80]. Hypoxia further leads to the HIF-1α-dependent upregulation of methionine adenosyltransferase 2A, which is an enzyme that synthesizes the methyl donor, SAM [81].

e) Cigarette smoking is also associated with a general inflammatory state and cell ratios in the body [82-85], which could also alter DNA methylation.

These large effects of cigarette smoking on genome-wide DNA methylation could further influence post-translational modifications possibly Immunoglobulin G glycosylation [86].

METHODOLOGICAL CONSIDERATIONS

Measurement of genome-wide methylation

The human genome contains more than 28 million CpGs [87], 70-80% of which are methylated [88]. However, the Illumina 450K arrays only measures 1.7% of the CpGs in the genome. The CpGs are unevenly distributed across the human genome. Methylation in promoters and enhancers create the most functional impact. Among the gene regions for CpGs included on this array, most (41%) of them are present at gene promoters, 31% at gene bodies, 35% at 3'UTR regions and 3% at intergenic regions. For the enhancer regions, that are not present for a large part on the Illumina 450k arrays, but are newly incorporated as 333,265 new CpG sites identified by the ENCODE [89, 90] and FANTOM5 [91] projects, in the recent Illumina 850k EPIC array [92]. It is possible that Hcy effects on DNA methylation may be specific to other sites and regions in the genome that are not measured by the Illumina 450k arrays, for example, enhancers. Therefore, future studies in this field can use the recent Illumina 850k EPIC arrays which has similar advantages as Illumina 450k arrays and not only contains >90% of the original content from Illumina 450K arrays but also includes high coverage from enhancer regions. Also, when calculating and analyzing genomic DNA methylation, i.e., the median of CpGs per sample using this Illumina 450k array data, one needs to keep in mind that the results might be biased due to limited genome coverage.

DATA-ANALYSIS

Normalization

A divergence between the β-values of the Infinium type I and type II probe assays was noticed previously wherein the β-values retrieved from the type II assay were of smaller range, less accurate and less stable as compared to the β-values retrieved from type I assay [93, 94]. In order to overcome this bias, several correction methods were proposed to rescale the type II β-values as per the type I β-values [93, 94] with each method having its own pros and cons. In the RS, we investigated the performance of these methods and although they all performed in a satisfactory manner, we decided to use DASEN [95] or quantile normalization from CPACOR workflow (separately done for Type 1 and Type 2 assays) [96] which performs *between samples* correction. The association analysis performs regression across samples per probe, and therefore it is important to perform *between samples* correction per probe instead of *across probes* correction per sample.

Differentially methylated regions

In the analysis from this thesis, we used the method Comb-p [97, 98] to estimate methylation differences across regions. This software was developed considering the

uneven distribution of CpGs on Illumina 450k arrays. It uses nominal p-values as input in a sliding window and also takes into account the correlation between CpGs associated with these p-values. This method uses only p-values as input, not only from single EWASs but also from meta-analyses. In the results described in Chapter 4 and Chapter 7, we found 68, 74 and 29 DMRs to be significantly associated with Hcy levels, dietary folate and dietary vitamin B12 intakes, respectively. The number of DMRs were much more than the observed DMPs. In addition, top DMR findings did not reflect a strong association for one or more single CpGs that were part of the identified DMR. Therefore, although DMR analysis on array-data is used in an increasing number of manuscripts, a standard method is yet to be identified. Since the design of the array is not optimal for DMR-analysis (large "gaps" between CpGs), simply "adding-up" results from two nearby CpGs seems too simple. A better option might be to include predefined regions based on databases with functional annotation of the genome, such as ENCODE. However, we still know very little about actual correlation of CpGs in different circumstances. All this, makes the "pure statistical" collapsing of different CpGs doubtful and results from these analysis should be carefully interpreted.

IMPLICATIONS AND FUTURE DIRECTIONS

Smoking-associated methylation changes showed relatively larger differences as compared to the factors that are more directly related to methylation-metabolism such as Hcy, Hcy-associated variants or B-vitamin intake. First, validation or replication is needed for the identified Hcy- , folate intake- and vitamin B12- associated DMPs. If true, they can be used as potential biomarkers in methyl-group metabolism related pathologies and for identification of new pathways and biological mechanisms associated with it. In addition, the smoking-associated DMPs could serve as sensitive markers for lifetime exposure to cigarette smoking. Although these DMPs need to have good sensitivity and specificity to be used as biomarkers for diagnostics, their use in diagnosis could serve many purposes [99, 100].

Since SAM is a major methyl donor and SAH is a potent inhibitor of methyl-transferase reactions, the ratio of SAM/SAH is thought to reflect the cellular methylation capacity [5]. Therefore, coming back to the possibilities discussed before as to why we found limited findings, measurements of SAM and SAH would be interesting to investigate methyl-group metabolism related methylation changes. However, SAM is an intrinsically unstable molecule [101], which needs controlled preanalytical conditions. Therefore, it is a challenging task to measure SAM or SAM:SAH ratio in population-based studies, and very few studies have studied this up to now [5-8].

Inorder to identify true effects of intermediate or severe HHcy i.e. more extreme dysregulated methyl-group metabolism, a population with a broader range of Hcy, i.e., individuals with severe HHcy or with folate/vitamin B12 deficiency should be

recruited. Furthermore, since dietary data is prone to measurement errors, measurements of plasma or cellular levels of vitamin B12 and folate should be conducted to confirm the role of these B-vitamins on DNA methylation. Tissues other than blood need to be investigated to find specific effects and more homogenous cell types should be studied for accuracy. Choosing the right tissue depends on the clinical interest. However, tissues such as heart, liver, kidney and brain are not readily available and are difficult to obtain in a substantial number of samples. Therefore, use of surrogate tissues such as blood or buccal cells can be improved by understanding the relationship between surrogate tissues and target tissues in an independent data set using more information, such as from correlated CpG sites [102]. Although methylation levels are largely conserved across tissues, it should be kept in mind that such prediction accuracy could vary across different tissue types, populations, phenotypes, diseases, genotypes and environmental exposure [102].

Intracellular Hcy, SAM and SAH

It could be that the effects of Hcy are seen at an intracellular level instead of plasma level, since methyl-group metabolism effectively takes place intracellularly. It was previously shown that plasma Hcy and intracellular SAH but not intracellular SAM were associated with lymphocyte DNA hypomethylation [5]. Moreover, heterozygous cystathionine β-synthase (CBS +/−) mice given a methyl-deficient diet for 24 weeks, showed that tissue-specific alterations in global DNA methylation levels of liver, kidney, brain and testes occurred not due to decreased intracellular SAM concentrations of these tissues alone, but due to increased intracellular SAH with or without combination with decreased intracellular SAM [103]. Brain and testes showed most sensitivity to alterations in SAH and DNA methylation, as compared with liver and kidney which can be mainly due to low SAH hydrolase activity and lack of complete transsulfuration pathway in these tissues [103]. Since for most methyl-group metabolites, plasma concentrations showed poor reflection of intracellular concentrations [9], future studies should measure intracellular Hcy, SAM and SAH concentrations and investigate their effect on genome-wide DNA methylation. This could be the reason why we did not find large differences of plasma homocysteine on genome-wide DNA methylation.

Hydroxymethylation

Hydroxymethylation is another kind of chemical modification which occurs from 5-methylcytosine to form 5-hydroxymethylcytosine (5hmC) catalyzed by ten-eleven translocation (TET) enzymes [104]. 5-hmC are known to play roles in gene expression, mammalian development, transcriptional regulation and human diseases [105]. It could be that the role of Hcy, B vitamin diet intake, other methyl-group metabolism factors and smoking is specific to some 5-hydroxycytosines. Like other techniques that are based on the bisulfite treatment, the Illumina 450k array is primarily incapable to

differentiate 5-methylcytosines from 5-hmC because both are resistant to bisulfite treatment [106-108]. Moreover, due to the low abundance of 5hmC as compared to 5-methylcytosines, it is technically more challenging to measure 5hmC [108]. Essentially, 5hmC can be measured by digestion- and antibody-based techniques [109] and LC/MS-MS [110, 111] or by TET-assisted bisulfite treatment followed by 450k arrays or sequencing.

Mendelian randomization

The relationship of Hcy, vitamin B12 and folate with DNA methylation can be subject to substantial bias, given the strong relationship between several lifestyle factors, diseases and methyl-group factors. To circumvent this bias, genetic factors determining the methyl-group factors as an instrument can be used to study the relationship between these factors and methylation by Mendelian Randomization (MR). This would eliminate the effects that are possibly caused by measurement errors, confounding and reverse causality. However, in Chapter 5 we performed a MR study of *MTHFR* 677C>T or genetic risk score (GRS). If a formal MR-study is conducted, the used SNPs or so called instruments need to satisfy 3 basic assumptions [112, 113]. One, the instrument should be associated with our exposure, Hcy. Two, the instrument should not associate with any confounder of the exposure-outcome association. Three, the instrument should not affect our outcome, DNA methylation, except through Hcy. This condition is also referred to as pleiotropy. Although all three assumptions are satisfied in the case of the *MTHFR* 677C>T [114], the GRS model might violate assumption three [112, 113]. Recent research in MR studies have developed several methods that can be used in the future to correct and estimate the correct causal estimates in case of pleiotropy [115-117].

Furthermore, the relationship between Hcy and DNA methylation should be known, before MR can be performed [118]. This relationship is only known in studies until now where DNA methylation is measured at a global level. Therefore, for the *MTHFR* 677C>T variant, we did not perform a complete MR because in Chapter 4 we did not find large effects of Hcy on site-specific DNA methylation, and therefore, we could not estimate the causal effect in this case. However, we performed a genetic study which focused on the association of genetically defined elevated Hcy levels with DNA methylation. The results were in line with the results described in Chapter 4, where subtle changes in genome-wide DNA methylation was observed.

Furthermore, there could be application of the MR technique for other exposures such as cigarette smoking, which has shown a larger effect on genome-wide DNA methylation as described in Chapter 8. A recent study of 822 individuals has already evaluated the previously found associations of cigarette smoking with 5 DMPs, using the MR study [119]. Using the SNPs rs4074134 for current smoking and rs9920506 for ever smoking as instruments, they found causal effect on

hypomethylation of two of the 5 DMPs of the *F2RL3* and *GPR15* genes. Future studies can conduct similar studies on the much larger list of the cigarette smoking associated DMPs that we identified in Chapter 8.

REFERENCES

1. Kang SS, Wong PW, Malinow MR. Hyperhomocyst(e)inemia as a risk factor for occlusive vascular disease. Annu Rev Nutr 1992;12:279-98.
2. Ingrosso D, Cimmino A, Perna AF, Masella L, De Santo NG, De Bonis ML, et al. Folate treatment and unbalanced methylation and changes of allelic expression induced by hyperhomocysteinaemia in patients with uraemia. Lancet 2003;361:1693-9.
3. Weiss N, Keller C, Hoffmann U, Loscalzo J. Endothelial dysfunction and atherothrombosis in mild hyperhomocysteinemia. Vasc Med 2002;7:227-39.
4. Melnyk S, Pogribna M, Pogribny IP, Yi P, James SJ. Measurement of plasma and intracellular S-adenosylmethionine and S-adenosylhomocysteine utilizing coulometric electrochemical detection: alterations with plasma homocysteine and pyridoxal 5'-phosphate concentrations. Clin Chem 2000;46:265-72.
5. Yi P, Melnyk S, Pogribna M, Pogribny IP, Hine RJ, James SJ. Increase in plasma homocysteine associated with parallel increases in plasma S-adenosylhomocysteine and lymphocyte DNA hypomethylation. J Biol Chem 2000;275:29318-23.
6. Castro R, Rivera I, Struys EA, Jansen EE, Ravasco P, Camilo ME, et al. Increased homocysteine and S-adenosylhomocysteine concentrations and DNA hypomethylation in vascular disease. Clin Chem 2003;49:1292-6.
7. Heil SG, Riksen NP, Boers GH, Smulders Y, Blom HJ. DNA methylation status is not impaired in treated cystathionine beta-synthase (CBS) deficient patients. Mol Genet Metab 2007;91:55-60.
8. Selley ML. A metabolic link between S-adenosylhomocysteine and polyunsaturated fatty acid metabolism in Alzheimer's disease. Neurobiol Aging 2007;28:1834-9.
9. Smith DE, Hornstra JM, Kok RM, Blom HJ, Smulders YM. Folic acid supplementation does not reduce intracellular homocysteine, and may disturb intracellular one-carbon metabolism. Clin Chem Lab Med 2013;51:1643-50.
10. Jung AY, Smulders Y, Verhoef P, Kok FJ, Blom H, Kok RM, et al. No effect of folic acid supplementation on global DNA methylation in men and women with moderately elevated homocysteine. PLoS One 2011;6:e24976.
11. Hughes CF, Ward M, Hoey L, McNulty H. Vitamin B12 and ageing: current issues and interaction with folate. Ann Clin Biochem 2013;50:315-29.
12. Institute of Medicine Standing Committee on the Scientific Evaluation of Dietary Reference I, its Panel on Folate OBV, Choline. 1998.
13. Green R, Allen LH, Bjorke-Monsen AL, Brito A, Gueant JL, Miller JW, et al. Vitamin B12 deficiency. Nat Rev Dis Primers 2017;3:17040.
14. Green R. Vitamin B12 deficiency from the perspective of a practicing hematologist. Blood 2017;129:2603-2611.
15. Gonseth S, Roy R, Houseman EA, de Smith AJ, Zhou M, Lee ST, et al. Periconceptional folate consumption is associated with neonatal DNA methylation modifications in neural crest regulatory and cancer development genes. Epigenetics 2015;10:1166-1176.
16. Joubert BR, den Dekker HT, Felix JF, Bohlin J, Ligthart S, Beckett E, et al. Maternal plasma folate impacts differential DNA methylation in an epigenome-wide meta-analysis of newborns. Nat Commun 2016;7:10577.
17. Waterland RA, Jirtle RL. Transposable elements: targets for early nutritional effects on epigenetic gene regulation. Mol Cell Biol 2003;23:5293-300.
18. Heijmans BT, Tobi EW, Stein AD, Putter H, Blauw GJ, Susser ES, et al. Persistent epigenetic differences associated with prenatal exposure to famine in humans. Proc Natl Acad Sci U S A 2008;105:17046-9.

19. Molloy AM, Kirke PN, Brody LC, Scott JM, Mills JL. Effects of folate and vitamin B12 deficiencies during pregnancy on fetal, infant, and child development. Food Nutr Bull 2008;29:S101-11; discussion S112-5.
20. McPartlin J, Halligan A, Scott JM, Darling M, Weir DG. Accelerated folate breakdown in pregnancy. Lancet 1993;341:148-9.
21. Higgins JR, Quinlivan EP, McPartlin J, Scott JM, Weir DG, Darling MR. The relationship between increased folate catabolism and the increased requirement for folate in pregnancy. Bjog 2000;107:1149-54.
22. Ronn T, Ling C. DNA methylation as a diagnostic and therapeutic target in the battle against Type 2 diabetes. Epigenomics 2015;7:451-60.
23. Bansal A, Pinney SE. DNA methylation and its role in the pathogenesis of diabetes. Pediatr Diabetes 2017;18:167-177.
24. Willer CJ, Bonnycastle LL, Conneely KN, Duren WL, Jackson AU, Scott LJ, et al. Screening of 134 single nucleotide polymorphisms (SNPs) previously associated with type 2 diabetes replicates association with 12 SNPs in nine genes. Diabetes 2007;56:256-64.
25. Peden JF, Farrall M. Thirty-five common variants for coronary artery disease: the fruits of much collaborative labour. Hum Mol Genet 2011;20:R198-205.
26. Adams JN, Raffield LM, Freedman BI, Langefeld CD, Ng MC, Carr JJ, et al. Analysis of common and coding variants with cardiovascular disease in the Diabetes Heart Study. Cardiovasc Diabetol 2014;13:77.
27. Fu Y, Dominissini D, Rechavi G, He C. Gene expression regulation mediated through reversible m(6)A RNA methylation. Nat Rev Genet 2014;15:293-306.
28. Batista PJ. The RNA Modification N6-methyladenosine and Its Implications in Human Disease. Genomics Proteomics Bioinformatics 2017;15:154-163.
29. Chandola U, Das R, Panda B. Role of the N6-methyladenosine RNA mark in gene regulation and its implications on development and disease. Brief Funct Genomics 2015;14:169-79.
30. Yang X, Yang Y, Sun BF, Chen YS, Xu JW, Lai WY, et al. 5-methylcytosine promotes mRNA export - NSUN2 as the methyltransferase and ALYREF as an m5C reader. Cell Res 2017;27:606-625.
31. Clarke S. Protein methylation. Curr Opin Cell Biol 1993;5:977-83.
32. Yang XD, Lamb A, Chen LF. Methylation, a new epigenetic mark for protein stability. Epigenetics 2009;4:429-33.
33. Bedford MT, Richard S. Arginine methylation an emerging regulator of protein function. Mol Cell 2005;18:263-72.
34. Esse R, Florindo C, Imbard A, Rocha MS, de Vriese AS, Smulders YM, et al. Global protein and histone arginine methylation are affected in a tissue-specific manner in a rat model of diet-induced hyperhomocysteinemia. Biochim Biophys Acta 2013;1832:1708-14.
35. Esse R, Rocha MS, Barroso M, Florindo C, Teerlink T, Kok RM, et al. Protein arginine methylation is more prone to inhibition by S-adenosylhomocysteine than DNA methylation in vascular endothelial cells. PLoS One 2013;8:e55483.
36. Lokk K, Modhukur V, Rajashekar B, Martens K, Magi R, Kolde R, et al. DNA methylome profiling of human tissues identifies global and tissue-specific methylation patterns. Genome Biol 2014;15:r54.
37. Ghosh S, Yates AJ, Fruhwald MC, Miecznikowski JC, Plass C, Smiraglia D. Tissue specific DNA methylation of CpG islands in normal human adult somatic tissues distinguishes neural from non-neural tissues. Epigenetics 2010;5:527-38.
38. Rakyan VK, Down TA, Thorne NP, Flicek P, Kulesha E, Graf S, et al. An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs). Genome Res 2008;18:1518-29.

39. Slieker RC, Bos SD, Goeman JJ, Bovee JV, Talens RP, van der Breggen R, et al. Identification and systematic annotation of tissue-specific differentially methylated regions using the Illumina 450k array. Epigenetics Chromatin 2013;6:26.
40. Devlin AM, Bottiglieri T, Domann FE, Lentz SR. Tissue-specific changes in H19 methylation and expression in mice with hyperhomocysteinemia. J Biol Chem 2005;280:25506-11.
41. Choumenkovitch SF, Selhub J, Bagley PJ, Maeda N, Nadeau MR, Smith DE, Choi SW. In the cystathionine beta-synthase knockout mouse, elevations in total plasma homocysteine increase tissue S-adenosylhomocysteine, but responses of S-adenosylmethionine and DNA methylation are tissue specific. J Nutr 2002;132:2157-60.
42. Glier MB, Ngai YF, Sulistyoningrum DC, Aleliunas RE, Bottiglieri T, Devlin AM. Tissue-specific relationship of S-adenosylhomocysteine with allele-specific H19/Igf2 methylation and imprinting in mice with hyperhomocysteinemia. Epigenetics 2013;8:44-53.
43. Zhou J, Sears RL, Xing X, Zhang B, Li D, Rockweiler NB, et al. Tissue-specific DNA methylation is conserved across human, mouse, and rat, and driven by primary sequence conservation. BMC Genomics 2017;18:724.
44. Lowe R, Gemma C, Beyan H, Hawa MI, Bazeos A, Leslie RD, et al. Buccals are likely to be a more informative surrogate tissue than blood for epigenome-wide association studies. Epigenetics 2013;8:445-54.
45. Varley KE, Gertz J, Bowling KM, Parker SL, Reddy TE, Pauli-Behn F, et al. Dynamic DNA methylation across diverse human cell lines and tissues. Genome Res 2013;23:555-67.
46. Lowe R, Slodkowicz G, Goldman N, Rakyan VK. The human blood DNA methylome displays a highly distinctive profile compared with other somatic tissues. Epigenetics 2015;10:274-81.
47. Jiang R, Jones MJ, Chen E, Neumann SM, Fraser HB, Miller GE, Kobor MS. Discordance of DNA methylation variance between two accessible human tissues. Sci Rep 2015;5:8257.
48. Walton E, Hass J, Liu J, Roffman JL, Bernardoni F, Roessner V, et al. Correspondence of DNA Methylation Between Blood and Brain Tissue and Its Application to Schizophrenia Research. Schizophr Bull 2016;42:406-14.
49. Hatchwell E, Greally JM. The potential role of epigenomic dysregulation in complex human disease. Trends Genet 2007;23:588-95.
50. Li Y, Zhu J, Tian G, Li N, Li Q, Ye M, et al. The DNA methylome of human peripheral blood mononuclear cells. PLoS Biol 2010;8:e1000533.
51. Michels KB, Binder AM, Dedeurwaerder S, Epstein CB, Greally JM, Gut I, et al. Recommendations for the design and analysis of epigenome-wide association studies. Nat Methods 2013;10:949-55.
52. Reinius LE, Acevedo N, Joerink M, Pershagen G, Dahlen SE, Greco D, et al. Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. PLoS One 2012;7:e41361.
53. Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. Genome Biol 2014;15:R31.
54. Lam LL, Emberly E, Fraser HB, Neumann SM, Chen E, Miller GE, Kobor MS. Factors underlying variable DNA methylation in a human community cohort. Proc Natl Acad Sci U S A 2012;109 Suppl 2:17253-60.
55. Lowe R, Rakyan VK. Correcting for cell-type composition bias in epigenome-wide association studies. Genome Med 2014;6:23.
56. Houseman EA, Kim S, Kelsey KT, Wiencke JK. DNA Methylation in Whole Blood: Uses and Challenges. Curr Environ Health Rep 2015;2:145-54.

57. Liang L, Cookson WO. Grasping nettles: cellular heterogeneity and other confounders in epigenome-wide association studies. Hum Mol Genet 2014;23:R83-8.
58. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics 2012;13:86.
59. Jones MJ, Islam SA, Edgar RD, Kobor MS. Adjusting for Cell Type Composition in DNA Methylation Data Using a Regression-Based Approach. Methods Mol Biol 2017;1589:99-106.
60. Chen S, Wu P, Zhou L, Shen Y, Li Y, Song H. Relationship between increase of serum homocysteine caused by smoking and oxidative damage in elderly patients with cardiovascular disease. Int J Clin Exp Med 2015;8:4446-54.
61. Haj Mouhamed D, Ezzaher A, Neffati F, Douki W, Najjar MF. Effect of cigarette smoking on plasma homocysteine concentrations. Clin Chem Lab Med 2011;49:479-83.
62. Vardavas CI, Linardakis MK, Hatzis CM, Malliaraki N, Saris WH, Kafatos AG. Smoking status in relation to serum folate and dietary vitamin intake. Tob Induc Dis 2008;4:8.
63. e Silva Ade S, da Mota MP. Effects of physical activity and training programs on plasma homocysteine levels: a systematic review. Amino Acids 2014;46:1795-804.
64. Kim YN, Hwang JH, Cho YO. The effects of exercise training and acute exercise duration on plasma folate and vitamin B12. Nutr Res Pract 2016;10:161-6.
65. Urgert R, van Vliet T, Zock PL, Katan MB. Heavy coffee consumption and plasma homocysteine: a randomized controlled trial in healthy volunteers. Am J Clin Nutr 2000;72:1107-10.
66. Nygard O, Refsum H, Ueland PM, Stensvold I, Nordrehaug JE, Kvale G, Vollset SE. Coffee consumption and plasma total homocysteine: The Hordaland Homocysteine Study. Am J Clin Nutr 1997;65:136-43.
67. Ulvik A, Vollset SE, Hoff G, Ueland PM. Coffee consumption and circulating B-vitamins in healthy middle-aged men and women. Clin Chem 2008;54:1489-96.
68. Gibson A, Woodside JV, Young IS, Sharpe PC, Mercer C, Patterson CC, et al. Alcohol increases homocysteine and reduces B vitamin concentration in healthy male volunteers--a randomized, crossover intervention study. Qjm 2008;101:881-7.
69. Bleich S, Bleich K, Kropp S, Bittermann HJ, Degner D, Sperling W, et al. Moderate alcohol consumption in social drinkers raises plasma homocysteine levels: a contradiction to the 'French Paradox'? Alcohol Alcohol 2001;36:189-92.
70. Lee KW, Pausova Z. Cigarette smoking and DNA methylation. Front Genet 2013;4:132.
71. Huang J, Okuka M, Lu W, Tsibris JC, McLean MP, Keefe DL, Liu L. Telomere shortening and DNA damage of embryonic stem cells induced by cigarette smoke. Reprod Toxicol 2013;35:89-95.
72. Mortusewicz O, Schermelleh L, Walter J, Cardoso MC, Leonhardt H. Recruitment of DNA methyltransferase I to DNA repair sites. Proc Natl Acad Sci U S A 2005;102:8905-9.
73. Cuozzo C, Porcellini A, Angrisano T, Morano A, Lee B, Di Pardo A, et al. DNA damage, homology-directed repair, and DNA methylation. PLoS Genet 2007;3:e110.
74. Lee EW, D'Alonzo GE. Cigarette smoking, nicotine addiction, and its pharmacologic treatment. Arch Intern Med 1993;153:34-48.
75. Shen JX, Yakel JL. Nicotinic acetylcholine receptor-mediated calcium signaling in the nervous system. Acta Pharmacol Sin 2009;30:673-80.
76. Satta R, Maloku E, Zhubi A, Pibiri F, Hajos M, Costa E, Guidotti A. Nicotine decreases DNA methyltransferase 1 expression and glutamic acid decarboxylase 67 promoter methylation in GABAergic interneurons. Proc Natl Acad Sci U S A 2008;105:16356-61.
77. Mercer BA, Wallace AM, Brinckerhoff CE, D'Armiento JM. Identification of a cigarette smoke-responsive region in the distal MMP-1 promoter. Am J Respir Cell Mol Biol 2009;40:4-12.

78. Di YP, Zhao J, Harper R. Cigarette smoke induces MUC5AC protein expression through the activation of Sp1. J Biol Chem 2012;287:27948-58.
79. Kadonaga JT, Carner KR, Masiarz FR, Tjian R. Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. Cell 1987;51:1079-90.
80. Olson KR. Carbon monoxide poisoning: mechanisms, presentation, and controversies in management. J Emerg Med 1984;1:233-43.
81. Liu Q, Liu L, Zhao Y, Zhang J, Wang D, Chen J, et al. Hypoxia induces genomic DNA demethylation through the activation of HIF-1alpha and transcriptional upregulation of MAT2A in hepatoma cells. Mol Cancer Ther 2011;10:1113-23.
82. Levitzky YS, Guo CY, Rong J, Larson MG, Walter RE, Keaney JF, Jr., et al. Relation of smoking status to a panel of inflammatory markers: the framingham offspring. Atherosclerosis 2008;201:217-24.
83. Lee J, Taneja V, Vassallo R. Cigarette smoking and inflammation: cellular and molecular mechanisms. J Dent Res 2012;91:142-9.
84. Saetta M, Turato G, Facchini FM, Corbino L, Lucchini RE, Casoni G, et al. Inflammatory cells in the bronchial glands of smokers with chronic bronchitis. Am J Respir Crit Care Med 1997;156:1633-9.
85. van der Vaart H, Postma DS, Timens W, ten Hacken NH. Acute effects of cigarette smoke on inflammation and oxidative stress: a review. Thorax 2004;59:713-21.
86. Wahl A, Kasela S, Carnero-Montoro E, van Iterson M, Stambuk J, Sharma S, et al. IgG glycosylation and DNA methylation are interconnected with smoking. Biochim Biophys Acta 2018;1862:637-648.
87. Stirzaker C, Taberlay PC, Statham AL, Clark SJ. Mining cancer methylomes: prospects and challenges. Trends Genet 2014;30:75-84.
88. Bird A. DNA methylation patterns and epigenetic memory. Genes Dev 2002;16:6-21.
89. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature 2012;489:57-74.
90. Siggens L, Ekwall K. Epigenetics, chromatin and genome organization: recent advances from the ENCODE project. J Intern Med 2014;276:201-14.
91. Lizio M, Harshbarger J, Shimoji H, Severin J, Kasukawa T, Sahin S, et al. Gateways to the FANTOM5 promoter level mammalian expression atlas. Genome Biol 2015;16:22.
92. Moran S, Arribas C, Esteller M. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. Epigenomics 2016;8:389-99.
93. Dedeurwaerder S, Defrance M, Calonne E, Denis H, Sotiriou C, Fuks F. Evaluation of the Infinium Methylation 450K technology. Epigenomics 2011;3:771-84.
94. Touleimat N, Tost J. Complete pipeline for Infinium((R)) Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. Epigenomics 2012;4:325-41.
95. Pidsley R, CC YW, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. BMC Genomics 2013;14:293.
96. Lehne B, Drong AW, Loh M, Zhang W, Scott WR, Tan ST, et al. A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. Genome Biol 2015;16:37.
97. Kechris KJ, Biehs B, Kornberg TB. Generalizing moving averages for tiling arrays using combined p-value statistics. Stat Appl Genet Mol Biol 2010;9:Article29.
98. Pedersen BS, Schwartz DA, Yang IV, Kechris KJ. Comb-p: software for combining, analyzing, grouping and correcting spatially correlated P-values. Bioinformatics 2012;28:2986-8.
99. Mikeska T, Bock C, Do H, Dobrovic A. DNA methylation biomarkers in cancer: progress towards clinical implementation. Expert Rev Mol Diagn 2012;12:473-87.

100. Leygo C, Williams M, Jin HC, Chan MWY, Chu WK, Grusch M, Cheng YY. DNA Methylation as a Noninvasive Epigenetic Biomarker for the Detection of Cancer. Dis Markers 2017;2017:3726595.
101. Hao X, Huang Y, Qiu M, Yin C, Ren H, Gan H, et al. Immunoassay of S-adenosylmethionine and S-adenosylhomocysteine: the methylation index as a biomarker for disease and health status. BMC Res Notes 2016;9:498.
102. Ma B, Wilker EH, Willis-Owen SA, Byun HM, Wong KC, Motta V, et al. Predicting DNA methylation level across human tissues. Nucleic Acids Res 2014;42:3515-28.
103. Caudill MA, Wang JC, Melnyk S, Pogribny IP, Jernigan S, Collins MD, et al. Intracellular S-adenosylhomocysteine concentrations predict global DNA hypomethylation in tissues of methyl-deficient cystathionine beta-synthase heterozygous mice. J Nutr 2001;131:2811-8.
104. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 2009;324:930-5.
105. Richa R, Sinha RP. Hydroxymethylation of DNA: an epigenetic marker. Excli J 2014;13:592-610.
106. Jin SG, Kadam S, Pfeifer GP. Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. Nucleic Acids Res 2010;38:e125.
107. Nestor C, Ruzov A, Meehan R, Dunican D. Enzymatic approaches and bisulfite sequencing cannot distinguish between 5-methylcytosine and 5-hydroxymethylcytosine in DNA. Biotechniques 2010;48:317-9.
108. Yong WS, Hsu FM, Chen PY. Profiling genome-wide DNA methylation. Epigenetics Chromatin 2016;9:26.
109. Kurdyukov S, Bullock M. DNA Methylation Analysis: Choosing the Right Method. Biology (Basel) 2016;5.
110. Song L, James SR, Kazim L, Karpf AR. Specific method for the determination of genomic DNA methylation by liquid chromatography-electrospray ionization tandem mass spectrometry. Anal Chem 2005;77:504-10.
111. Le T, Kim KP, Fan G, Faull KF. A sensitive mass spectrometry method for simultaneous quantification of DNA methylation and hydroxymethylation levels in biological samples. Anal Biochem 2011;412:203-9.
112. Swanson SA, Hernan MA. Commentary: how to report instrumental variable analyses (suggestions welcome). Epidemiology 2013;24:370-4.
113. Martens EP, Pestman WR, de Boer A, Belitser SV, Klungel OH. Instrumental variables: application and limitations. Epidemiology 2006;17:260-7.
114. van Meurs JB, Pare G, Schwartz SM, Hazra A, Tanaka T, Vermeulen SH, et al. Common genetic loci influencing plasma homocysteine concentrations and their effect on risk of coronary artery disease. Am J Clin Nutr 2013;98:668-76.
115. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. Int J Epidemiol 2017;46:1985-1998.
116. van Kippersluis H, Rietveld CA. Pleiotropy-robust Mendelian randomization. Int J Epidemiol 2017.
117. Schmidt AF, Dudbridge F. Mendelian randomization with Egger pleiotropy correction and weakly informative Bayesian priors. Int J Epidemiol 2017.
118. Dekkers KF, van Iterson M, Slieker RC, Moed MH, Bonder MJ, van Galen M, et al. Blood lipids influence DNA methylation in circulating cells. Genome Biol 2016;17:138.
119. Jhun MA, Smith JA, Ware EB, Kardia SLR, Mosley TH, Jr., Turner ST, et al. Modeling the Causal Role of DNA Methylation in the Association Between Cigarette Smoking and

Inflammation in African Americans: A 2-Step Epigenetic Mendelian Randomization Study. Am J Epidemiol 2017;186:1149-1158.

CHAPTER 10: SUMMARY

Homocysteine and DNA methylation

Homocysteine (Hcy) is a key metabolite of the methyl-group metabolism, elevated levels of which in a medical condition called hyperhomocysteinemia (HHcy) are associated with several disorders such as cardiovascular diseases and osteoporotic fractures. Being a byproduct of the methyltransferase reaction, Hcy's role in alteration of DNA methylation levels is thought as one of the underlying mechanisms of Hcy-associated disorders. Several studies have investigated the association between Hcy and DNA methylation in both animal models and human subjects. We reviewed this literature in **Chapter 2** and found that in animal models, both diet- and genetically-induced elevated Hcy showed alteration in DNA methylation indicating Hcy as a keyplayer. However in human subjects, elevated Hcy did not show consistent alterations in DNA methylation. These contradictory findings in humans might be explained by the differences in tissue types, small sample sizes, pathophysiological effects of the underlying disease, genetic differences between humans, limited studies and the fact that DNA methylation was measured on a global scale.

To follow up on our understanding of this review, we conducted a epigenome-wide association study (EWAS) of Hcy in **Chapter 4** to investigate the association between plasma Hcy and DNA methylation in leukocytes of 700 individuals of the RS. We increased our sample size up to 2,035 individuals by collaborating with other cohorts and conducted a meta-analysis of the individual EWAS results per cohort. We found that plasma Hcy was associated with subtle changes in DNA methylation, wherein we identified some novel epigenetic loci; three differentially methylated sites (DMPs) and 68 differentially methylated regions (DMRs) associated with Hcy.

Given the strong relationship between several lifestyle factors, diseases and Hcy, the relationship between Hcy and DNA methylation can be subject to substantial bias possibly caused by measurement errors, confounding and reverse causality. Therefore we eliminated these effects in **Chapter 5** by using genetically defined elevated Hcy, i.e. the most consistent *MTHFR* 677C>T variant and the combined weighted genetic risk score (GRS) of 18 previously studied Hcy-associated variants, to test whether genetically defined elevated Hcy levels are associated with DNA methylation changes in leukocytes of 9,894 individuals from 12 cohorts. Except for the cis-DMPs (< 1Mb upstream or downstream of the SNP), we did not observe widespread changes in DNA methylation; mainly one trans-DMPs (< 1Mb upstream or downstream of the SNP) associated with the *MTHFR* 677C>T variant and five trans-CpGs associated with the GRS.

Lastly in **Chapter 6**, we conducted an interaction study of plasma Hcy and *MTHFR* 677C>T variant to investigate their combined effect on genome-wide DNA methylation in leukocytes of 1280 individuals from 2 cohorts. We identified 13 DMPs

associated with [Hcy x *MTHFR* 677C>T]. This included four cis-DMPs at the *AGTRAP-MTHFR-NPPA/B* gene cluster on chromosome 1, which were also found in Chapter 5, indicating that this association was driven by the *MTHFR* 677C>T variant. We did not see overlaps of any other DMPs between our analysis of plasma Hcy, *MTHFR* 677C>T/GRS and [Hcy x *MTHFR* 677C>T] in Chapter 4, 5 and 6, respectively.

Nutrition, Lifestyle and DNA methylation

The metabolites and co-enzymes that participate in the methyl-group metabolism are methionine, choline, betaine and B-vitamins, namely folate, vitamin B2, B6 and B12. Folate and vitamin B12 are widely studied co-enzymes and are essential micronutrients involved in the donation of methyl groups in cellular metabolism, deficiencies of which could impair the remethylation pathway and lead to elevated Hcy. Therefore, general nutrition intake including the status of these vitamins could play a vital role in DNA methylation. We conducted a systematic review in **Chapter 3** to study the relation between micro- and macro- nutrients and DNA methylation in disease-free individuals across the life course. Micronutrients included B-vitamins, where in most studies, we found that folate but not vitamin B12 was associated with DNA methylation. However, the directions of these associations were inconsistent across studies. This was attributed to differences across life course and nutrient and methylation tissue type.

To study the effect of the nutritional folate and vitamin B12 intake on a large epigenome-wide scale, we conducted meta-analyses in leukocytes of 5,841 participants from EWASs of 10 cohorts in **Chapter 7**, where we identified some novel epigenetic loci; mainly 6 novel DMPs and 74 DMRs significantly associated with dietary folate intake, and 29 DMRs significantly associated with dietary vitamin B12 intake. Although both dietary intake and cellular levels of folate are important determinants of plasma Hcy, we did not see overlaps of the six folate intake-associated DMPs with our analysis of plasma Hcy, *MTHFR* 677C>T/GRS and [Hcy x *MTHFR* 677C>T] in Chapter 4, 5 and 6, respectively.

In addition to the one carbon metabolism and nutritional factors, smoking is a crucial lifestyle factor for disorders such as respiratory diseases, cardiovascular diseases, cancer and reproductive outcomes has been previously associated with epigenome-wide DNA methylation. However, large-scale meta-analysis studies with increased sample sizes have not been conducted yet. In **Chapter 8**, we focused on association between cigarette smoking as a lifestyle factor and DNA methylation in leukocytes, assessed in 15,907 individuals (2,433 current, 6,518 former, and 6,956 never smokers) from EWASs of 16 cohorts. We identified 2,623 DMPs at bonferroni threshold $P<1\times10^{-7}$ (18,760 DMPs at false discovery rate (FDR) <0.05) when comparing current versus never smokers, and 185 DMPs at bonferroni threshold $P<1\times10^{-7}$ (2,623 DMPs at FDR <0.05) when comparing former versus never smokers, indicating a pattern of persistently altered methylation, with attenuation, after smoking cessation.

HOOFDSTUK 10: SAMENVATTING

Homocysteïne en DNA-methylatie

Homocysteïne (Hcy) is een belangrijke metaboliet in het metabolisme van methylgroepen. Verhoogde concentraties van homocysteïne in het bloed, in medische terminologie hyperhomocysteïnemie (HHcy) genoemd, zijn geassocieerd met verschillende stoornissen zoals hart- en vaatziekten en osteoporotische fracturen. Omdat Hcy een bijproduct is van de DNA-methyltransferasereactie, wordt een verandering in DNA-methylering beschouwd als één van de mogelijke pathologische mechanismen bij Hcy-geassocieerde aandoeningen. Verschillende studies hebben de associatie tussen Hcy en DNA-methylering bij zowel diermodellen als bij mensen onderzocht. Wij hebben deze literatuur in **Hoofdstuk 2** samengevat en daarin aanwijzingen gevonden dat in diermodellen zowel dieet- als genetisch geïnduceerde HHcy veranderingen in DNA-methylering tot gevolg hebben die aan Hcy worden toegeschreven. Bij de mens vertoonde verhoogd Hcy echter geen consistente veranderingen in DNA-methylering. Deze tegenstrijdige bevindingen kunnen worden verklaard door verschillen in onderzochte weefseltypes, steekproefgroottes, pathofysiologische effecten van de onderliggende ziekte, genetische verschillen tussen mensen, opzet van de studies en het feit dat meestal globale DNA-methylering werd gemeten.

In aansluiting op dit overzicht van de literatuur hebben wij in **Hoofdstuk 4** een epigenoom-brede associatiestudie (EWAS) van Hcy uitgevoerd om de associatie tussen plasma Hcy en DNA-methylering in leukocyten van 700 individuen van de Rotterdam Studie (RS) te onderzoeken. Wij hebben onze steekproefomvang verder verhoogd tot 2.035 personen door samen te werken met andere cohorten en hebben een meta-analyse uitgevoerd van de individuele EWAS-resultaten per cohort. Wij vonden dat plasma-Hcy geassocieerd was met subtiele veranderingen in DNA-methylering, waarbij een aantal nieuwe epigenetische loci werd geïdentificeerd: drie differentieel gemethyleerde posities (DMPs) en 68 differentieel gemethyleerde regio's (DMRs) geassocieerd met Hcy.

Gezien de sterke relatie tussen verschillende leefstijlfactoren, ziekten en Hcy, kan de relatie tussen Hcy en DNA-methylering onderworpen zijn aan substantiële bias, veroorzaakt door meetfouten, verstorende factoren of omgekeerde causaliteit. In **Hoofdstuk 5** hebben wij getracht deze factoren te elimineren door gebruik te maken van genetisch vastgelegde HHcy, dwz van de meest consistente *MTHFR* 677C>T variant en de gecombineerde gewogen genetische risicoscore (GRS) van 18 eerder bestudeerde Hcy-geassocieerde varianten, om te testen of deze genetisch gedefinieerde verhoogde Hcy niveaus zijn geassocieerd met DNA-methyleringsveranderingen in de leukocyten van 9.894 individuen uit 12 cohorten. Met uitzondering van cis-DMP's (<1Mb stroomopwaarts of stroomafwaarts van de SNP),

hebben wij geen wijdverspreide veranderingen in DNA-methylering waargenomen; wel één trans-DMP's (<1Mb stroomopwaarts of stroomafwaarts van de SNP) geassocieerd met de *MTHFR* 677C>T variant en vijf trans-CpG's geassocieerd met de GRS.

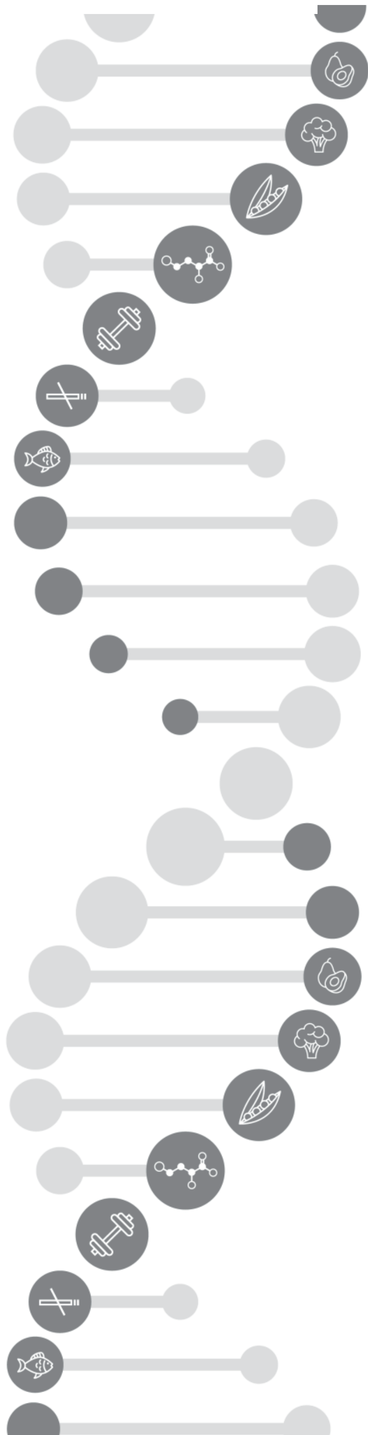
Ten slotte hebben we in **Hoofdstuk 6** een interactiestudie uitgevoerd van plasma Hcy en de *MTHFR* 677C>T variant om hun gecombineerde effect op genoom-brede DNA-methylering in leukocyten van 1280 individuen uit 2 cohorten te onderzoeken. Wij identificeerden 13 DMP's geassocieerd met [Hcy x *MTHFR* 677C>T]. Dit omvatte vier cis-DMP's in het *AGTRAP-MTHFR-NPPA*/*B* gencluster op chromosoom 1, die ook werden gevonden in hoofdstuk 5, en dit geeft aan dat deze associatie wordt aangedreven door de *MTHFR* 677C>T variant. We hebben geen overlappingen van andere DMP's gezien tussen onze analyses van plasma Hcy, *MTHFR* 677C>T/GRS en [Hcy x *MTHFR* 677C>T] in respectievelijk de hoofdstukken 4, 5 en 6.

Voeding, levensstijl en DNA-methylatie

De metabolieten en co-enzymen die deelnemen aan het methyl-groep-metabolisme zijn methionine, choline, betaïne en de B-vitamines folaat, vitamine B2, B6 en B12. Folaat en vitamine B12 zijn veel bestudeerde micronutriënten die in het celmetabolisme als coenzym betrokken zijn bij de overdracht van methylgroepen , en waarvan tekorten de remethylatie-route kunnen verstoren en zo leiden tot verhoogd Hcy. Daarom zou algemene voedingsinname inclusief de status van deze vitaminen, een vitale rol kunnen spelen bij DNA-methylering. Wij hebben in **Hoofdstuk 3** een systematische review uitgevoerd om de relatie tussen micro- en macronutriënten en DNA-methylering in ziektevrije individuen gedurende de levensloop te bestuderen. B-vitaminen zijn micronutriënten, waarvan we in de meeste onderzoeken vonden dat folaat wel maar vitamine B12 niet geassocieerd was met DNA-methylering. De richting van deze associaties was in verschillende studies echter niet consistent. Dit werd toegeschreven aan verschillen in levensloop, type voeding en het type weefsel onderzocht op methylering.

Om het effect van de folaat en vitamine B12 inname uit voeding op een grote, epigenoom-brede schaal te bestuderen, hebben we in **Hoofdstuk 7** een meta-analyses uitgevoerd in leukocyten van 5.841 deelnemers van EWASs van 10 cohorten , en daarin enkele nieuwe epigenetische loci geïdentificeerd: zes nieuwe DMP's en 74 DMR's die significant geassocieerd zijn met foliumzuurinname via het dieet en 29 DMR's die significant geassocieerd zijn met de inname van vitamine B12. Hoewel zowel de inname via de voeding als de cellulaire niveaus van folaat belangrijke determinanten zijn van plasma Hcy, hebben we geen overlappingen gezien van de zes foliumzuurinname-geassocieerde DMP's met onze analyses van plasma Hcy, *MTHFR* 677C>T/GRS en [Hcy x *MTHFR* 677C>T] in de hoofdstukken 4, 5 en 6.

Naast het methylgroep-metabolisme en voedingsfactoren, is roken een cruciale leefstijlfactor die wordt gerelateerd aan aandoeningen van de luchtwegen, hart- en vaatziekten, kanker en zwangerschap én eerder is geassocieerd met veranderingen in epigenoom-brede DNA-methylering. Er zijn echter nog geen resultaten van grootschalige meta-analyseonderzoeken met toegenomen steekproefgroottes bekend. In **Hoofdstuk 8** hebben we ons gericht op de associatie tussen het roken van sigaretten als leefstijlfactor en DNA-methylering in leukocyten, door middel van EWASs bij 15.907 personen uit 16 cohorten (2.433 actieve, 6.518 voormalige en 6.956 nooit-rokers). Wij identificeerden 2.623 DMP's bij Bonferroni-drempel P <1 × 10-7 (18.760 DMP's bij fout-positieven ratio (FDR) <0,05) bij vergelijking van actieve versus nooit-rokers, en 185 DMP's bij Bonferroni-drempelwaarde P <1 x 10-7 (2.623 DMP's bij FDR <0,05) bij vergelijking van voormalige versus nooit-rokers, hetgeen wijst op een patroon van aanhoudend veranderde maar wel verzwakkende methylering na stoppen met roken.



PART E

Appendices

LIST OF PUBLICATIONS

IN THIS THESIS

1. **Mandaviya PR**, Stolk L, Heil SG. Homocysteine and DNA methylation: a review of animal and human literature. *Mol Genet Metab 2014;113:243-52.*
2. **Mandaviya PR***, Aïssi D*, Dekkers KF*, Joehanes R*, Kasela S*, Truong V*, Stolk L, Heemst DV, Ikram MA, Lindemans J, Slagboom PE, Trégouët DA, Uitterlinden AG, Wei C, Wells P, Gagnon F[#], van Greevenbroek MM[#], Heijmans BT[#], Milani L[#], Morange PE[#], van Meurs JB[#], Heil SG[#]; BIOS Consortium. Homocysteine levels associate with subtle changes in leukocyte DNA methylation: an epigenome-wide analysis. *Epigenomics 2017;9:1403-1422.*
3. **Mandaviya PR***, Joehanes R*, Aïssi D*, Kühnel B*, Marioni RE*, Truong V*, Stolk L, Beekman M, Bonder MJ, Franke L, Gieger C, Huan T, Ikram MA, Kunze S, Liang L, Lindemans J, Liu C, McRae AF, Mendelson MM, Müller-Nurasyid M, Peters A, Slagboom PE, Starr JM, Trégouët DA, Uitterlinden AG, van Greevenbroek MMJ, van Heemst D, van Iterson M, Wells PS, Yao C, Deary IJ[#], Gagnon F[#], Heijmans BT[#], Levy D[#], Morange PE[#], Waldenberger M[#], Heil SG[#], van Meurs JBJ[#]; CHARGE Consortium Epigenetics group and BIOS Consortium. Genetically defined elevated homocysteine levels do not result in widespread changes of DNA methylation in leukocytes. *PLoS One 2017;12:e0182472.*
4. Joehanes R*, Just AC*, Marioni RE*, Pilling LC*, Reynolds LM*, **Mandaviya PR***, Guan W*, Xu T*, Elks CE*, Aslibekyan S*, Moreno-Macias H*, Smith JA*, Brody JA*, Dhingra R*, Yousefi P, Pankow JS, Kunze S, Shah SH, McRae AF, Lohman K, Sha J, Absher DM, Ferrucci L, Zhao W, Demerath EW, Bressler J, Grove ML, Huan T, Liu C, Mendelson MM, Yao C, Kiel DP, Peters A, Wang-Sattler R, Visscher PM, Wray NR, Starr JM, Ding J, Rodriguez CJ, Wareham NJ, Irvin MR, Zhi D, Barrdahl M, Vineis P, Ambatipudi S, Uitterlinden AG, Hofman A, Schwartz J, Colicino E, Hou L, Vokonas PS, Hernandez DG, Singleton AB, Bandinelli S, Turner ST, Ware EB, Smith AK, Klengel T, Binder EB, Psaty BM, Taylor KD, Gharib SA, Swenson BR, Liang L, DeMeo DL, O'Connor GT, Herceg Z, Ressler KJ, Conneely KN[#], Sotoodehnia N[#], Kardia SL[#], Melzer D[#], Baccarelli AA[#], van Meurs JB[#], Romieu I[#], Arnett DK[#], Ong KK[#], Liu Y[#], Waldenberger M[#], Deary IJ[#], Fornage M[#], Levy D[#], London SJ[#]. Epigenetic Signatures of Cigarette Smoking. *Circ Cardiovasc Genet 2016;9:436-447.*
5. Nash AJ*, **Mandaviya PR***, Dib M, Uitterlinden AG, van Meurs JBJ, Heil SG[#], Andrew T[#] and Ahmadi KR[#]. Interaction between plasma homocysteine and the *MTHFR c.677C>T* polymorphism is associated with site-specific changes in DNA methylation in humans. *Submitted for publication.*
6. **Mandaviya PR***, Joehanes R*, Brody J*, Castillo-Fernandez JE*, Dekkers KF*, Do AN*, Graff M*, Hänninen IK*, Tanaka T*, de Jonge EAL, Kiefte-de Jong JC, Absher DM, Aslibekyan S, de Rijke YB, Fornage M, Hernandez DG, Hurme MA, Ikram MA,

Jacques PF, Justice AE, Kiel DP, Lemaitre RN, Mendelson MM, Mikkilä V, Moore AZ, Pallister T, Raitakari OT, Schalkwijk CG, Sha J, Slagboom EPE, Smith CE, Stehouwer CDA, Tsai P, Uitterlinden AG, van der Kallen CJH, van Heemst DD, BIOS Consortium, CHARGE Consortium Epigenetics group, CHARGE Consortium Nutrition group, Arnett DK#, Bandinelli S#, Bell JT#, Heijmans BBT#, Lehtimäki T#, Levy D#, North KE#, Sotoodehnia N#, van Greevenbroek MMJ#, van Meurs JBJ#, Heil SG#. Association of dietary folate and vitamin B12 intake with genome-wide DNA methylation; a large scale epigenome-wide association analysis in 5,841 individuals. *Submitted for publication.*

7. Braun KVE*, **Mandaviya PR***, Franco OH, Nano J, Girschik C, Bramer WM, Muka T, Troup J, van Meurs JBJ, Heil SG, Voortman T. Nutrients and DNA methylation across the life course: a systematic review. *To be submitted.*

OTHER PUBLICATIONS

8. Karlsson Linnér R*, Marioni RE*, Rietveld CA*, Simpkin AJ, Davies NM, Watanabe K, Armstrong NJ, Auro K, Baumbach C, Bonder MJ, Buchwald J, Fiorito G, Ismail K, Iurato S, Joensuu A, Karell P, Kasela S, Lahti J, McRae AF, **Mandaviya PR**, Seppälä I, Wang Y, Baglietto L, Binder EB, Harris SE, Hodge AM, Horvath S, Hurme M, Johannesson M, Latvala A, Mather KA, Medland SE, Metspalu A, Milani L, Milne RL, Pattie A, Pedersen NL, Peters A, Polidoro S, Räikkönen K, Severi G, Starr JM, Stolk L, Waldenberger M; BIOS Consortium, Eriksson JG, Esko T, Franke L, Gieger C, Giles GG, Hägg S, Jousilahti P, Kaprio J, Kähönen M, Lehtimäki T, Martin NG, van Meurs JBC, Ollikainen M, Perola M, Posthuma D, Raitakari OT, Sachdev PS, Taskesen E, Uitterlinden AG, Vineis P, Wijmenga C, Wright MJ, Relton C, Davey Smith G, Deary IJ#, Koellinger PD#, Benjamin DJ#. An epigenome-wide association study meta-analysis of educational attainment. *Mol Psychiatry. 2017 Dec;22(12):1680-1690.*

9. Den Dekker HT, Burrows K, Felix JF, Salas LA, Nedeljkovic I, Yao J, Rifas-Shiman SL, Ruiz-Arenas C, DeMeo DL, Henderson AJ, Howe CG, Hivert M, Ikram MA, de Jongste JC, Lahousse L, **Mandaviya PR**, van Meurs JB, Pinart M, Stolk L, Sunyer J, Uitterlinden AG, Anto JM, Litonjua AA, Breton CV, Brusselle GG, Bustamante M, Davey Smith G, Relton CL, Jaddoe VWV, Duijts L. Newborn DNA-methylation, childhood lung function, and the risk of asthma and COPD across the life course. *Submitted for publication.*

10. Peters FS, Peeters AMA, **Mandaviya PR**, van Meurs JBJ, Hofland LJ, van de Wetering J, Betjes MGH, Baan CC, Boer K. Differentially methylated regions in T cells identify kidney transplant patients at risk for de novo skin cancer. *Submitted for publication.*

11. Lin H, Lunetta KL, Zhao Q, **Mandaviya PR**, Rong J, Benjamin EJ, Joehanes R, Levy D, van Meurs JBJ, Larson MG, Murabito JM. Whole Blood Gene Expression Associated with Clinical Biological Age. *Submitted for publication.*

12. Luijk R, Wu H, Ward-Caviness C, Hannon E, Carnero-Montoro E, Min J, **Mandaviya PR**, Mei H, van der Maarel S, Jansen R, Relton C, Mill J, Waldenberger M, Bell J, Franke L, ‘t Hoen PAC, Boomsma DI, van Duin CM, van Greevenbroek MMJ, Veldink JH, Wijmenga C, van Meurs JBJ[19], Daxinger L, Slagboom PE, van Zwet EW, Heijmans BT. Autosomal genetic loci are associated with variable X-chromosome inactivation. *Submitted for publication.*
13. Ince-Askan H, **Mandaviya PR**, Felix JF, van Meurs JBJ, Hazes JMW, Dolhain RJEM. Altered DNA methylation in children born to mothers with rheumatoid arthritis during pregnancy. *To be submitted.*
14. **Mandaviya PR**, van Rooij J, Stutvoet MD, Felix JF, van Dongen J, Janssen R, Franke L, ’t Hoen PAC, Claringbould AJ, Heijmans BT, van Meurs JBJ. Creating standardized association catalogs between the functional genome (epigenetics/RNA-expression) and BBMRI phenotypes: comparison of data analysis methodologies. *In preparation.*
15. Parmar P, Lowry E, Cugliari G, Suderman M, Wilson R, Wielscher M, Karhunen V, Teumer A, Lehne B, Milani L, de Klein N, Mishra PP, Melton PE, **Mandaviya PR**, Kasela S, Muka T, Zhang Y, Uitterlinden AG, Peters A, Geiger C, Anderson D, Boomsma DI, Grabeb HJ, Veldink JH, van Meurs JBJ, van den Berg L, Beilen L, Franke L, van Greevenbroek MMJ, Nauck M, Kahonen M, Hurme MA, Raitakari OT, Franco OH, Slagboom PE, van der Harst P, Kunze S, Felix SB, Zhang T, Chen W, Mori T, Heijmans BT, Nano J, Chambers JC, Fischer K, Huang R, Waldenberger M, Lehtimäki T, Rathmann W, Relton CL, Matullo G, Brenner H, Verweij N, Li S, Jarvelin M, Sebert S. A meta-analysis of maternal smoking *GFI1*-CpGs and cardio-metabolic phenotypes in adults. *In preparation.*
16. Wielscher M, Lehne B, Dehghan A, **Mandaviya PR**, van Meurs JBJ, Heijmans BT, BIOS Consortium, Boomsma DI, Slagboom PE, van Greevenbroek MMJ, Veldink JH, van den Berg L, Chambers J, Jarvelin MR, et al. Meta-analysis of genome-wide DNA methylation and serum CRP levels. *In preparation.*
17. The GoDMC consortium. 250,000 independent genetic influences on DNA methylation and the consequences of these perturbations: the GoDMC Consortium. *In preparation.*

*These authors contributed equally as first authors
#These authors contributed equally as last authors

PHD PORTFOLIO

Name: Pooja Rajendra Mandaviya
Erasmus MC Department: Internal Medicine
Clinical Chemistry
Research school: NIHES & MolMed
PhD period: August 2012 – May 2018
Promoters: Prof. Dr. André G. Uitterlinden,
Prof. Dr. Jan Lindemans
Supervisors: Dr. Joyce B.J. van Meurs
Dr. Sandra G. Heil

COURSES	Month, Year (Duration)	Duration	Workload (ECTS)
Research management	Oct 23 - Nov 13, 2014	2 days	1.0 ECTS
Biomedical English writing	Sep 9 - Oct 21, 2014	3 days	2.0 ECTS
Integrity in research	Jul 15, 2014	1 day	0.3 ECTS
Biostatistical methods: Basic principles (CC02)	Sep 23 - Oct 18, 2013	4 days	5.7 ECTS
Presenting Skills	Apr 29 - Jun 7, 2013	3 days	1.0 ECTS
Introduction Course on SPSS	Feb 19 - 21, 2013	3 days	0.8 ECTS
SNP Course	Nov 19 - 23, 2012	1 week	2.0 ECTS
			12.8 ECTS

CONGRESSES	Month, Year (Duration)	Presentation	Workload (Hours)
Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Investigator Meeting, *Rotterdam, Netherlands*	April 18-19, 2018	Oral & Poster	16 hours
Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Investigator Meeting, *Boston, United States*	September 27-28, 2017	Poster	16 hours
DUSRA (Dutch Society for Research on Aging), *Leiden, Netherlands*	May 19, 2017	Poster	8 hours
Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Investigator Meeting, *Charlottesville, United States*	September 27-29, 2016	Poster	24 hours
Biobank-based Integrative Omics Studies (BIOS) Symposium, *Amsterdam, Netherlands*	September 21, 2016	No	8 hours
4th International Vitamin Conference, *Copenhagen, Denmark*	May 25-27, 2016	Oral	24 hours
10th International Conference On One Carbon Metabolism, Vitamins B and Homocysteine, *Nancy, France*	July 7-11, 2015	Oral & Poster	40 hours

19th Molecular Medicine Day (MolMed), *Rotterdam, Netherlands*	March 19, 2015	Poster	8 hours
Science days 2015, *Antwerpen, Belgium*	January 8-9, 2015	Poster	16 hours
Dutch Society for Human Genetics (NVHG), *Arnhem, Netherlands*	October 2-3, 2014	Poster	16 hours
18th Molecular Medicine Day (MolMed), *Rotterdam, Netherlands*	February 20, 2014	Poster	8 hours
9th International Conference On Homocysteine and One-Carbon Metabolism, *Dublin, Ireland*	September 8-12, 2013	Poster	40 hours
2nd Annual Infinium Human Methylation 450k Array Workshop, *London, United Kingdom*	April 15, 2013	Poster	8 hours
17th Molecular Medicine Day (MolMed), *Rotterdam, Netherlands*	February 13, 2013	No	8 hours
Netherlands Consortium for Healthy Ageing (NCHA), *The Hague, Netherlands*	February 7-8, 2013	No	16 hours
			240 hours

OTHER RESEARCH/TEACHING ACTIVITIES	**Year**	**Workload (Hours/Months)**
Conducted quality control and pre-processing of:		
○ The Rotterdam Study Illumina 450k methylation datasets	2015 & 2017	3 months
○ An Illumina 450k methylation project titled: “Epigenetic changes in children born to mothers with Rheumatoid Arthritis” (Including analysis)	2017	2.5 months
○ An EPIC (850k) array project	2017	16 hours
○ The B-Vitamins for the PRevention Of Osteoporotic Fractures (BPROOF) study Illumina 450k methylation dataset	2016	3 months
Taught practical session on Bioinformatics tools for the Erasmus Summer programme “Genomics in Molecular Medicine” at Erasmus MC, Netherlands	2014 & 2016	12 hours
Referee for the journal of Clinical Biochemistry	2015	3 hours
Experience in writing an Erasmus MC grant	2014	16 hours
Assisted with homocysteine measurements	2013	32 hours

ABOUT THE AUTHOR

Pooja Rajendra Mandaviya was born in Porbandar, the coastal city of Gujarat, India, on 20th October, 1987. She completed high school at St. Joseph's institute and intermediate school at Murgaon Education Society Higher Secondary School, Goa university, Goa, India. She obtained her bachelors' degree in Biotechnology in May 2009 from Garden City College, Bangalore university, Bangalore, India, which brought together concepts of cell and molecular biology, genetics and other disciplines of life sciences. She obtained her masters' degree in Bioinformatics in July 2011 from the School of Life Sciences, Manipal University, Manipal, India. In addition to the concepts of cell and molecular biology, her masters' degree focused on biological databases, systems biology, information science, biostatistics and data analysis.

During her master, she pursued a 6 months research internship at the Bioinformatics department BiGCaT, Maastricht University, Maastricht, the Netherlands. Her research internship focused on manual and automated extension of microRNA target information on existing pathways for WikiPathways and Cytoscape visualization tools. Automated extension was done by implementing a Cytoscape plug-in written in Java programming. After her Masters' degree, she worked as a database curator for 6 months to read and extract information of the esophageal squamous cell carcinoma for a Not-For-Profit team, MBiome, which works for Bioinformatics solutions.

In August 2012, she started her PhD at the departments of Internal Medicine and Clinical Chemistry of the Erasmus University Medical Center, Rotterdam, the Netherlands, under the supervision of Dr. Sandra G. Heil and Dr. Joyce B.J. van Meurs. She worked on the Rotterdam Study, a large prospective cohort and conducted several epigenome-wide association analyses to investigate the associations of the key methyl-group metabolism factors such as homocysteine, *MTHFR* 677C>T, B-vitamin intakes, and other lifestyle factors such as cigarette smoking with site-specific genome-wide DNA methylation. She collaborated with many other research groups to bring together similar kind of data in a meta-analysis approach.

She is currently pursuing her research at the department of Internal Medicine, Erasmus University Medical Center, Rotterdam, the Netherlands, under the supervision of Dr. Joyce B.J. van Meurs.

WORDS OF GRATITUDE

I would like to first thank the five important people whose guidance and support has together made a truly great team in order to have me achieve my PhD.

To my promoters: Prof. dr. André Uitterlinden and Prof. dr. Jan Lindemans
André, during my interview process in May 2012, you had happily welcomed me in your group. Looking back to where I started, it has been a big learning curve for me. Your cheer and guidance during my presentations at work discussions and conferences have always filled me with motivation to do better. I am very grateful to you for this opportunity and your support in having me at our department for an extended period.

Jan, thank you for your huge support until the very end of my PhD. Discussions and ideas from you about the methyl group metabolism have guided me with reasoning and understanding different aspects. You have reviewed my thesis with so much attention to fine details. Thank you for being so actively available and being the driving force, even after your retirement, and congratulating me with success at every step.

To my co-promoters: Dr. Sandra G. Heil and Dr. Joyce van Meurs
Sandra, I honestly wouldn't have been so lucky if I didn’t have a supervisor like you. When I started my PhD, I had difficulty prioritizing things and lacked organization. Your perfectionism has helped me overcome my limitation in a way that it is now not only helping me in my professional life, but also my personal life. I must say I am not completely there yet, but I know I am on my way to it. You always had the time for me even if it wasn't so important. I have enjoyed having discussions with you. One of the qualities that I learnt from you is to think through thoroughly before making a decision.

Joyce, I have a little confession to make. I was very nervous of you initially because I perceived you as a very strict personality. Your critical comments on my work used to give me chills, until one day I decided that I will not pay attention to how I feel, but only focus on progressing my work. After travelling with you at the CHARGE conference, Charlottesville and working with you more, I realized that you are actually very warm hearted from the inside. Your guidance has helped me perform better and shaped our projects to the optimal level. Thank you for being so considerate, patient and supportive for me with everything and in having me here for an extended period.

To my former co-promoter: Dr. Lisette Stolk
Lisette, I would like to express my sincere gratitude for your supervision during the first two and a half years of my PhD. You helped me understand and have a good grip on the different analytical methods of the DNA methylation data. I really appreciate all

your contribution in helpful discussions, setting up analysis plans and helping me connect with other research groups for effective collaborations.

To the members of the doctoral committee

Prof. Arfan Ikram, Prof. Joost Gribnau and Prof. Caroline Relton, I am very happy that you were a part of the reading committee. Prof. Robert de Jonge, Prof. Bas Heijmans and Prof. Lisette de Groot, thank you for agreeing to serve on my doctoral committee. I truly appreciate the time and expertise of all the committee members in assessment of my thesis.

To my colleagues of the Department of Internal Medicine

Eline, I genuinely thank you for all your good administrative help during my PhD. You have been really kind. Marijn, I appreciate your continuous IT help and the nice lunches we had with the group. Cindy, Linda, Carolina, Jeroen and Marjolein (former), thank you for being super helpful whenever I was stuck with questions. I do look up to you guys. Carolina, I am grateful to you for your statistical help at so many times throughout my PhD. Jeroen, thank you for your kindness and patience whenever I needed more explanations. I thank Fernando and all other colleagues for their effective suggestions during my presentations. Long, I appreciate your help with my questions regarding Microsoft word. You are really good with figuring things out. Mila, Pascal, Michael and Joost, it has always been nice to conversate with you. Thanks for your ever friendliness. Kate, I really think you are funny and I have enjoyed sharing laughters with you. Thanks for being helpful always. Fjorda, you are truly a wonderful person. Thank you for your kindness and care. Enisa, thanks for the nice sweets that you always bring to our room. Jinluan, it has been so positive and fun having you in the room. Thanks for your care and feeding us with all the nice Chinese food. I also thank Guan, Jia-Lian (former), Natalia (former), Denise, Olja, Ingrid and Djawad for the nice conversations.

To my colleagues of the Department of Clinical Chemistry

Bertrand and Pieter, thank you for the insightful discussions during our epigenetics meetings. I thank Prof. dr. Yolanda B. de Rijke for her careful evaluation of the dietary B-vitamin EWAS manuscript. Samira, Pamela, Maja, Marianne, Ethan (former), Natanja, Helen, Rutchanna and Chérina, I have really enjoyed having you guys around. Thanks for the nice lunches, *mince pies*, *lemper*, plenty of *nussinis'*, support for my small marathon dream (smile), and all your good help and nice conversations. Richard, thanks for those nice walks and your IT help. I also thank Corine, Natascha and Evert for the nice lunches and conversations. Gigi and Iqbal, I have enjoyed the food you used to bring for me. Thanks for being so warm and caring. Cor, I thank you and your wife for the tour you gave us to Kinderdijk and surrounding area. It was such a great

experience by having a local friend to show us around. Chérina, words can't describe how thankful I am. Thanks for being like a mentor to me and being there always.

To my colleagues of the Department of Epidemiology
Janine, I am grateful for your quick helps and good input whenever I emailed you with questions. Jessica, Ester, Oscar, Trudy and Kim, thanks for the good discussions on the nutrition related projects. It was a wonderful experience working with you. Kim, I also think that you are an excellent review partner. We had a perfect grip together because we understood each other's OCDs. I also thank Natalie, Dina, Hata, Ashley, Jun, Shahzad, Ivana, Ajorla and Najaf for their friendly conversations whenever we crossed each other in the hallway. Najaf, thanks for your good advices and the useful tips you gave me when I was starting to write my thesis introduction and discussion.

To my collaborators of the BIOS consortium, CHARGE consortium and other groups
Prof. Bas Heijmans, I thank you for your valuable guidance and discussions during my PhD. It is a pleasure having you as a mentor. Koen Dekker, Maarten van Iterson and Marc Jan Bonder, thank you for your good contribution to our projects and your fast help to my emails. I am thankful to Dr. Caren E. Smith, Dr. Myriam Fornage, Dr. Marleen MJ van Greevenbroek, Dr. David-Alexandre Trégouët, Dr. France Gagnon, Dr. Daniel Levy, Dr. Roby Joehannes, Dr. Kourosh R Ahmadi, Dr. Toby Andrew, Dr. Paul F. Jacques, Dr. Diana van Heemst, Dr. Michael M. Mendelson and all other collaborators for their useful contributions, discussions and careful corrections on the manuscripts.

To my previous mentors
To the mentors of the department of BiGCaT, Maastricht University; Prof. Chris Evelo, Dr. Susan Steinbusch-Coort, Dr. Martina Kutmon, Dr. Thomas Kelder and Dr. Lars Eijssen. I am very glad to have done my Masters' project with you and I am grateful that you introduced this PhD to me. I thank my mentors of the Masters' training: Prashantha Hebbar, Dr. Prasad Burra and the one I connected through MBiome, Dr. Manoj Kashyap, for their valuable guidance and support before my PhD. I thank my mentors of school who were a great inspiration and are my well-wishers even today.

To my friends and family
I truly thank my friends and family friends for their support and presence in many ways and in different phases during my PhD and even before. I thank my parents, grandmother, brother, cousins and all other family members for their well-wishes, understanding and support. I mean much more than my words can express.

You can't connect the dots looking forward;
You can only connect them looking backwards.
(Steve Jobs)