Imperial College London

The impact of secular trends, dietary patterns and socioeconomic conditions on DNA methylation:

an epigenome-wide investigation

A dissertation by:

Christiana Demetriou

CID:00548483

Submitted in Fulfillment of the Requirements for the Degree of

Doctor of Philosophy (PhD)

Imperial College London, Department of Epidemiology and Biostatistics

Declaration of Originality

'I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or material which have been accepted for the award of any other degree or diploma at Imperial College or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at Imperial College or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.'

Signed: 

Date: 15th August, 2013

Copyright Declaration

‘The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.’

Acknowledgements

Undertaking this PhD has been a truly life-changing experience for me and it would not have been possible to do without the support and guidance that I received from many people.

I would like to first say a very big thank you to both my supervisors, Prof. Paolo Vineis and Dr. Kyriacos Kyriacou, for all their support and encouragement, during these long 3 years, both during my visits at Imperial College as well as while I remained in Cyprus. Without their guidance and constant feedback this PhD would not have been achievable.

Many thanks also to Dr James Flanagan who has devoted many hours meeting with me, sharing ideas, and giving feedback related to my PhD thesis and publications. Without his suggestions and guidance the quality of the results obtained would have been lacking.

Of course, this PhD would not have been possible without data to analyse, therefore a big thank you to Prof. Jia Chen and Dr. Silvia Polidoro for analysing and providing me with the data in this thesis.

A very special "thank you" also goes out to the members of both my departments at Imperial College as well as at the Cyprus Institute of Neurology and Genetics. Karin van Veldhoven thank you for being my link between Cyprus and London, for keeping me up to date with all the developments, for helping me and for venting with me during the long months of statistical uncertainty. Gianluca Campanella, thank you for all the hours you have put into pre-processing our data, creating the CH_3 package, and answering my un-intelligent questions. You are a treasure! Ioanna Neophytou, Maria Loizidou, Maria Kyratzi, Maria Kakkoura, Nasia Patsia, Sofia Neophytou, Giota Pirpa, Christina Flouri, and Emily Charalambidou, thank you for making the long days at the Institute enjoyable, for taking my mind off the stress, and for making me laugh my way through this PhD. Ioanna, Maria L., Maria K. and Andreas Hadjisavvas also thank you for all the time you have put into reading and providing feedback on my thesis and publications.

In addition, I gratefully acknowledge the funding I received from the Cyprus Civil Society and the FP7 project Linked2Safety during the years of my PhD.

Last but not least, I have no words through which to express my gratitude to my family, and especially my parents Chrystalla Demetriou and Andreas Demetriou, for supporting me in many ways through all the years of my education. Your upbringing has been my driving force and inspiration to become all that I could be. And of course a special thank you to the one who has been there through the good and the bad and who, at the end of the day, has made all of this worth it, Antonis Katsourides.

Abstract

Globalisation has resulted in several anthropometric, reproductive, and lifestyle factors and/or exposures exhibiting secular trends. These same factors and exposures have been implicated in modifying breast cancer risk and have been shown to influence DNA methylation in at least one publication. In this thesis, I conduct a preliminary and exploratory investigation to examine whether socioeconomic status, nutritional factors, weight and height, age at menarche, menopausal status, age at menopause, and estimated lifetime estrogen exposure (ELEE) – all factors displaying secular trends and influencing breast cancer risk – are associated with DNA methylation.

The LUminometric Methylation Assay (LUMA) was used to assess global methylation in 376 female subjects from the EPIC cohort and the Infinium HumanMethylation 450 BeadChip, was used to assess genome-wide methylation in 667 male and female subjects from the EPIC Italy subcohort. Genome wide, probe-type specific, CpG locus specific, clustering and single enrichment analyses were performed.

Total caloric intake, folate intake, total dietary methyl intake, and age at menarche were found to be associated with global methylation using LUMA. In addition, the 450k BeadChip results suggest that education, household density, highest household occupational position, total caloric intake, folate intake, dietary methyl intake, adherence to the Italian Mediterranean diet, body mass index, age at menopause and ELEE influence methylation patterns, but the power of these investigations did not permit identification of biologically relevant, differentially methylated CpG sites.

The present findings indicate that DNA methylation may be one way through which these exposures may become biologically embedded and influence disease risk, and especially breast cancer risk. Mechanisms involving epigenetic changes provide a possible mechanism though which even early life exposures can affect adult disease risk. However, more detailed and more powered epidemiological study designs and tools need to be implemented in order to confirm these associations and also examine causality.

Table of Contents

Declaration of Originality 2
Copyright Declaration 3
Acknowledgements 4
Abstract 5
Table of Contents 6
Table of Figures 9
List of Tables 11
List of Abbreviations 12
List of Thesis-Allied Publications 16
1 Introduction – Background 17
1.1 Globalization and Secular Trends 18
1.1.1 Socioeconomic Status 19
1.1.2 Nutrition and Dietary Habits 21
1.1.3 Weight, BMI and Obesity 23
1.1.4 Age at Menarche / Pubertal Development 27
1.1.5 Age at Menopause 30
1.1.6 Height 31
1.2 Secular Changes and Health Effects 33
1.2.1 Socioeconomic Status Health Effects 33
1.2.2 Nutritional Health Effects 35
1.2.3 Overweight and Obesity Health Effects 37
1.2.4 Age at Menarche and Age at Menopause Health Effects 38
1.2.5 Height Health Effects 39
1.2.6 Secular Changes and Breast Cancer 40
1.2.7 Biological Embedding of Trends and Exposures 43
1.3 Breast Cancer and Epigenetics: a narrative review 45
1.3.1 The Epigenetics of Breast Cancer 45
1.3.2 Epigenetics as Biomarker for Early Detection/Diagnosis 53
1.3.3 Epigenetics as Biomarker for Environmentally Induced Risk 55
1.3.4 Epigenetics as Prognostic Tool for Response to Treatment 56
1.3.5 Epigenetics as Potential Cancer Treatment 57
1.4 DNA Methylation and Features Undergoing Secular Changes 58
1.4.1 Socioeconomic Status and DNA Methylation 58
1.4.2 Nutrition and DNA methylation 63
1.4.3 Overweight/Obesity and DNA Methylation 70

1.4.4 Age at Menarche/Menopause and DNA methylation 79
1.4.5 Height and DNA methylation 81
1.5 Hypothesis and Study Aims 83
2 Subjects and Methods 84
2.1 Subjects and Methods for Global Methylation Analyses 84
2.1.1 Study Participants 84
2.1.2 Methods for Global Methylation Measurement 84
2.1.3 Statistical Analyses 86
2.2 Subjects and Methods for Genome Wide Methylation Analyses 90
2.2.1 Study Participants: 90
2.2.2 Methods for Genome wide Methylation measurement 90
2.2.3 Statistical Analyses 93
3 Results 111
3.1 Global Methylation (LUMA) 111
3.2 Genome Wide Methylation (Illumina 450K BeadChip) 123
3.2.1 Normalisation and Pre-processing 123
3.2.2 Whole Genome Methylation analysis 128
3.2.3 Genome Wide locus by locus Methylation Analyses 132
3.2.4 Clustering 144
3.2.5 Principal Component Analysis 148
3.2.6 Single Enrichment Analysis 150
4 Discussion 153
4.1 Discussion of Results 154
4.1.1 Socioeconomic Status 154
4.1.2 Nutrition 157
4.1.3 Anthropometry 160
4.1.4 Age at Menarche – Menopausal Status – Age at Menopause – ELEE 162
4.2 General Remarks 166
4.3 Difficulties in Establishing Causality in Environment-Epigenome Investigations 169
4.4 Limitations of the Investigation 172
4.5 Future Directions 174
5 Conclusions 178
References 179
APPENDIX A 203
APPENDIX B 216

APPENDIX C 239
APPENDIX D 246
Table 1 – Diet and DNA Methylation Review: Studies on diet and gene specific methylation 246
Table 2 – Diet and DNA Methylation Review: Studies on diet and global methylation 252
APPENDIX E 257
APPENDIX F 300
APPENDIX G 329

Table of Figures

Figure 1 –Trends in Socioeconomic Status. 19
Figure 2 – Unemployment rates by gender, EU, seasonally adjusted, January 2000 - February 2013. 20
Figure 3 – Association of obesity prevalence and age among women of the Diogenes cohort. 23
Figure 4 – Prevalence of female overweight and obesity, adults ≥ 20 years, by country and year, 1980–2008. 25
Figure 5 – Obesity among US children by age, 1988-1994 through to 2009-2010. 26
Figure 6 – Changes in prevalence of overweight and obesity per year. 26
Figure 7 – Mean age at menarche of women born between 1915 and 1964. 27
Figure 8 – Average height (cm) at time for conscription per country from 1880 to 2006. 31
Figure 9 – Mortality among children younger than five years of age and percentage of deprived households (lacking three or more essential items) in selected countries in the WHO European Region. 33
Figure 10 – Model of the pathways through which SES influences health. 34
Figure 11 – Folate in one-carbon metabolism. 63
Figure 12 – The predominant effect of ethanol on one-carbon metabolism appears to be the inhibition of methionine synthase. 65
Figure 13 – Literature review on dietary factors and DNA methylation: Flow Chart 67
Figure 14 – Investigations in studies of the association between measures of folate exposure and global DNA methylation. 69
Figure 15 – Luminometric output 85
Figure 16 – Infinium Methylation Assay scheme. 92
Figure 17 – Peak-based correction method for improving Infinium II (type II) data which are less accurate and reproducible than Infinium I (type I) data. 95
Figure 18 – Diagnostic plot for dye bias correction following background subtraction. 125
Figure 19 – Plot showing BMIQ normalization of type 1 (type I) and type 2 (type II) probes for a random sample (subject #100). 125
Figure 20 – Removal of batch effects in the entire dataset (n=667, 57 chips) using ComBat. 126
Figure 21 – Removal of batch effects in the “females only” dataset (n=485, 53 chips) using ComBat. 127
Figure 22 – Description of the 293,873 retained probes for whole genome methylation analyses. 129
Figure 23 – Description of the 444,278 retained probes for locus by locus analyses. 132
Figure 24 – Volcano plots illustrating the negative logarithm of the P-value from the regression analysis of each probe against the difference in medians between early life socioeconomic status exposure groups for that probe. 133
Figure 25 – Graphical representations of P-value density distributions in locus by locus analyses of early life socioeconomic status variables. 134
Figure 26 – Volcano plots illustrating the negative logarithm of the P-value from the regression analysis of each CpG site against the difference in medians between adult socioeconomic status exposure groups for that probe. 135
Figure 27 – Graphical representations of P-value distributions in locus by locus analyses of adult socioeconomic status variables. 136
Figure 28 – A. Volcano plots illustrating the negative logarithm of the P-value from the regression analysis of each CpG site against the difference in medians between anthropometric variable exposure groups for that probe. B. Graphical representations of P-value distributions in locus by locus analyses of anthropometric variables. 137
Figure 29 – Volcano plots illustrating the negative logarithm of the P-value from the regression analysis of each CpG site against the difference in medians between dietary variable exposure groups for that probe. 138

Figure 30 – Graphical representations of P-value distributions in locus by locus analyses of nutritional variables..... 140
Figure 31 – Volcano plots illustrating the negative logarithm of the P-value from the regression analysis of each CpG site against the difference in medians between reproductive variable exposure groups for that probe. 142
Figure 32 – Graphical representations of P-value distributions in locus by locus analyses of reproductive variables..... 143
Figure 33 – Heatmaps displaying hierarchical clustering of subjects based on the 1000 most variable probes, using Euclidean distance with average linkage. 144
Figure 34 – Consensus clustering output graphics for the entire dataset (male and female subjects)..... 145
Figure 35 – Consensus clustering output graphics for the 'females only' dataset. 146
Figure 36 – Functional classification of genes differentially methylated between household density exposure categories. 150
Figure 37 – Functional annotation clustering of genes differentially methylated between household density exposure categories. 151
Figure 38 – Functional classification of genes differentially methylated between estimated lifetime estrogen exposure categories. 152
Figure 39 – Functional annotation clustering of genes differentially methylated between estimated lifetime estrogen exposure categories. 152
Figure 40 – The complementarity of the exposome 169
Figure 41 – Two-step epigenetic Mendelian randomisation: applying the principle of Mendelian randomisation to DNA methylation as an intermediate phenotype. 176
Figure 42 – The Meet-in-the-Middle approach: an epidemiological tool to reinforce causal inference in epigenetic epidemiology. 177

List of Tables

Table 1 – Studies investigating the association between Socioeconomic Status and DNA methylation..... 60
Table 2 – Studies investigating the association between overweight/obesity and DNA methylation 73
Table 3 – Studies investigating the association between age at menarche and/or age at menopause and DNA methylation 80
Table 4 – Studies investigating the association between height and DNA methylation..... 82
Table 5 – Exposures undergoing secular trends and DNA methylation. 83
Table 6 – Calculation of a methyl intake dietary score 88
Table 7 – Variables examined for each exposure of interest 102
Table 8 – Variables examined and confounders adjusted for in linear regression models with methylation as continuous outcome..... 104
Table 9 – Demographic, anthropometric, lifestyle, and reproductive characteristics for subjects with LUMA measurements..... 111
Table 10 – Anthropometric and lifestyle variables in healthy subjects with respect to LUMA methylation quartiles..... 113
Table 11 – Association of highest level of education attained and birth during WWII with above/below median global methylation in a logistic regression model..... 115
Table 12 – Median and inter quartile range of daily intakes of specific food groups by percent global methylation group. P-values for Mann-Whitney Test..... 116
Table 13 – Median and inter quartile range of daily intakes of specific nutrients and energy by percent global methylation group. P-values for Mann-Whitney Test. 117
Table 14 – Associations between alcohol, folate and energy daily intakes with percent global methylation in a logistic regression model adjusting for relevant confounders 118
Table 15 – Association between adherence to the Mediterranean diet and above/below median global methylation in a logistic regression model 118
Table 16 – Association between dietary methyl intake and above/below median global methylation in a logistic regression model..... 119
Table 17 – Anthropometric variables and association with global methylation in a logistic regression model 120
Table 18 – Reproductive variables and association with global methylation in a logistic regression model 122
Table 19 – Demographic, anthropometric, lifestyle, and reproductive characteristics for subjects with Genome-Wide Methylation measurements 123
Table 20 – Number of probes retained for whole genome methylation and clustering analyses, following removal of compromising probes 128
Table 21 – Associations between each exposure variable and average methylation across different functional genomic distribution and CpG Island context probe classes. 131
Table 22 – Significantly differentially methylated CpGs between age at menarche and Estimated Lifetime Oestrogen Exposure (ELEE) categories..... 141
Table 23 – Association between Consensus Clustering derived clusters and exposure category for each exposure variable of interest..... 147
Table 24 – Associations between Principal Components and Subject Variables 149
Table 25 – Summary of Results per exposure examined 153

List of Abbreviations

Abbreviation	*In full*
5hmC	5-hydroxymethylCytosine
5mC	5-methylCytosine
ABCA1	ATP-binding cassette, sub-family A (ABC1), member 1
ACACB	Acetyl-CoA Carboxylase Beta
ACCN2	Acid-sensing (proton-gated) ion channel 1
ACOX1	Acyl-CoA Oxidase 1, palmitoyl
ALPL	Alkaline Phosphatase
ANOVA	Analysis of Variance
APC	Adenomatous Polyposis Coli
APOC1	Apolipoprotein C-I
AQP9	Aquaporin 9
ARF	ADP-ribosylation factor
ATM	Ataxia Telangiectasia Mutated
ATP10A	ATPase, class V, type 10A
BBD	Benign Breast Disease
BMAL1	Aryl Hydrocarbon Receptor Nuclear Translocator-like
BMI	Body Mass Index
BMIQ	Beta Mixture Quantile Normalisation
BRCA1	Breast Cancer 1 gene
BRCA2	Breast Cancer 2 gene
CASP10_P2	Caspase 10, apoptosis-related cysteine peptidase
CC	Consensus Clustering
CCND2	Cyclin D2
CD44	CD44 molecule
CDF	Cumulative Distribution Function
CDH1	Cadherin-1
CDK10	Cyclin Dependent Kinase 10
CDK1C	Cyclin-Dependent Kinase 1C
CDKN2B	Cyclin Dependent Kinase Inhibitor 2B
CHD	Coronary Heart Disease
CIMP	CpG Island Methylator Phenotype
CKMT1B	Creatine Kinase, mitochondrial 1B
CLOCK	Clock Circadian Regulator
CpG	Cytosine-phosphodiester bond-Guanine
CRH	Corticotropin Releasing Hormone
CVD	CardioVascular Disease
DAPK	Death-Associated Protein Kinase 1
dCTPs	Deoxycytidine triphosphates
DMR	Differentially Methylated Region
DNA	DeoxyriboNucleic Acid
DNMT	DNA Methyl Transferase
dNTP	Deoxyribonucleotide
DUSP22	Dual Specificity Phosphatase 22
EDCs	Endocrine Disrupting Chemicals
EGFR	Epidermal Growth Factor Receptor
ELEE	Estimated Lifetime oEstrogen Exposure
eNOS	Endothelial Nitric Oxide Synthetase

EPHA1	EPH receptor A1
EPIC	European Prospective Investigation into Cancer and Nutrition
ER	Estrogen Receptor
ESR-1	Estrogen Receptor-1
EU	European Union
EXOC5	Exocyst Complex Component 5
FDR	False Discovery Rate
FFTP	First Full Term Pregnancy
FGD	Fuctional Genomic Distribution
FTO	Fat mass and Obesity associated
GLMs	Generalized Linear Models
GNAS A/B	GNAS complex locus A/B
GNASAS	GNAS antisense RNA 1
GNP	Gross National Product
GRB10	Growth Factor Receptor-bound protein 10
GSTP1	Glutathione S-Transferase Pi 1
H3K9	Histone 3 Lysine 9
H3K20	Histone 3 Lysine 20
H3K27	Histone 3 Lysine 27
HC	Hierarchical Clustering
HDACs	Histone DeAcetylases
HDMs	Histone DeMethylases
HIC1	Hypermethylated In Cancer
HIN1	Secretoglobin, family 3A, member 1
HIPK3	Homeodomain Interacting Protein Kinase 3
HLA_BOB3	Major Histocompatibility Complex, class II, DO beta
HMTs	Histone MethylTransferases
HPA	Hypothalamic-Pituitary-Adrenal axis
HRT	Hormone Replacement Therapy
ICA	Independent Component Analysis
IDC	Invasive Ductal Carcinoma
IFR5	Interferon Regulatory Factor - 5
IGF	Insulin-like Growth Factor
IL10	Interleukin 10
IL6	Interleukin 6 (interferon, beta 2)
IL8	Interleukin 8
ILC	Invasive Lobular Carcinoma
IMI	Italian Mediterranean Index
INPP5J	Inositol Polyphosphate-5-phosphatase J
INSIGF	Insulin- Insulin-Like Growth Factor 2 Read-Through
ISVA	Independent Surrogate Variable Analysis
KCNQ1OT1	KCNQ1 opposite strand/antisense transcript 1 (non-protein coding)
KLF9	Kruppel-like factor 9
KNN	K-Nearest Neighbour
LEP	Leptin
LINE1	Long Interspersed Nuclear Element - 1
LOH	Loss of Heterozygosity
LUMA	Luminometric Methylation Assay
MAGE	Melanoma Antigen Gene
MCHR1	Melanin-concentrating Hormone Receptor 1

MDR1	Multidrug Resistance 1 gene
MeCP2	Methyl CpG Binding Protein
MEG3	Maternally Expressed 3 (non-protein coding)
methylTHF	methyl TetraHydroFolate
MGMT	O6 methylguanine DNA methyltransferase
MINT	Spen Homolog, transcriptional regulator
MLH1	MutL homolog 1
MMP9	Matrix MetalloPeptidase 9
MPL	MyeloProliferative Leukemia virus oncogene
MRPL41	Mitochondrial Ribosomal Protein L41
MSH2	DNA mismatch repair protein
MTHFR	methyl TetraHydroFolate Reductase
MYH2	Myosin, heavy chain 2
NCDs	Non Communicable Diseases
NID2	Nidogen 2 (osteonidogen)
NR3C1	Nuclear Receptor Subfamily 3, group C, member 1 (glucocorticoid receptor)
OC	Oral Contraceptives
OR	Odds Ratio
OXA1l	Oxidase (Cytochrome c) Assembly 1-like
P13KCD	Catalytic D Polypeptide
p16 / CDKN2A	Cyclin Dependent Kinase Inhibitor 2A
PARP	Poly ADP Ribose Polymerase
PBL	Peripheral Blood Leucocytes
PC	Principal Component
PCA	Principal Component Analysis
PcGs	Polycomb Group
PDHA-1	Pyruvate Dehydrogenase (lipoamide) alpha 1
PDP1	Pyruvate Dehyrogenase Phosphatase catalytic subunit 1
PER2	Period Circadian Clock 2
PITX2	Paired-like Homeodomain transcription factor 2
PM20D1	Peptidase M20 Domain Containing 1
POMC	Preproopiomelanocortin
PPP1R3C	Protein Phosphatase 1, regulatory subunit 3C
PR	Progesterone Receptor
PRKAA1	Protein Kinase, AMP-activated, alpha 1 catalytic subunit
PRKG1	Protein Kinase, cGMP-dependent, type I
PTP4A3M	Protein Tyrosine Phosphatase type IVA, member 3
Q	Quartile
RARβ2	Retinoic Acid Receptor beta-2
RASSF1A	Ras association domain containing protein 1
RFC5	Replication Factor C (activator 1) 5
RNA	RiboNucleic Acid
ROS	Reactive Oxygen Species
RUNX3	Runt-Related Transcription Factor 3
RXRA	Retinoid X Receptor - α
SAdoMet	S-adenosylmethionine
Sat2	Satellite-2
SEA	Single Enrichment Analysis
SEP	SocioEconomic Position
SES	SocioEconomic Status

SFRP1	Secreted Frizzled-Related Protein 1
SINUCLEIN γ	Gamma-Synuclein
SLC16A1	Solute Carrier Family 16 (monocarboxylate transporter), member 1
SMAD6	SMAD family member 6
SNP	Single Nucleotide Polymorphisms
Socs1	Suppressor of cytokine signalling 1
SOD	Superoxide Dismutase-1
SQN	Subset Quantile Normalisation
SVA	Surrogate Variable Analysis
SVD	Single Value Decomposition
SWAN	Subset Quantile Within Array Normalisation
TIMP3	TIMP metallopeptidase inhibitor 3
TNF	Tumour Necrosis Factor
TNNI3	Troponin I type 3
TNNT1	Troponin T type 1
TP53	Tumour Protein 53
TRIM3	Tripartite Motif containing 3
TRβ1	Thyroid Hormone Receptor β1
UBASH3A	Ubiquitin Associated and SH3 domain containing A
uPA	Plasminogen Activator
US	United States
WBCs	White Blood Cells
WFA	Weight-for-Age
WHO	World Health Organisation
WNT1	Wingless-type MMTV Integration Site family, member 1
WWII	World War II

List of Thesis-Allied Publications

- Buckland G, Travier N, Agudo A, Fonseca-Nunes A, Navarro C, Lagiou P, **Demetriou CA** et al. Olive oil intake and breast cancer risk in the Mediterranean countries of the European Prospective Investigation into Cancer and Nutrition study. *Int J Cancer*. 2012 Mar 6.
- **Demetriou CA**, Hadjisavvas A, Loizidou MA, Loucaides G, Neophytou I, Sieri S, Kakouri E, Middleton N, Vineis P, Kyriacou K. The Mediterranean dietary pattern and breast cancer risk in Greek-Cypriot women: a case-control study. *BMC Cancer*. 2012 Mar 23;12:113.
- Herceg Z, Lambert M-P, van Veldhoven K, **Demetriou CA**, Vineis P, Smith M, et al. Towards incorporating epigenetic mechanisms into carcinogen identification and evaluation. IARC. Carcinogenesis. First published online: June 7, 2013
- **Demetriou CA**, Hadjisavvas A, Loizidou MA, Vineis P, Kyriacou K. The Mediterranean Diet and Breast Cancer. In Preedy VR (Ed.), *The Mediterranean Diet: An Evidence-Based Approach To Disease Prevention* (1st Edition*)*. London: Academic Press. Expected publication: February 2015
- **Demetriou CA**, Chen J, Polidoro S, van Veldhoven K, Cuenin C, Campanella G et al. Methylome Analysis Reveals Epigenetic Changes Associated with Menarcheal Age. Accepted for publication PLoS ONE.

1 Introduction – Background

In the wake of the era of westernisation and very rapid changes in socioeconomic circumstances, especially in low and middle income nations, several lifestyle and anthropometric variables are displaying secular trends, mostly affected or driven by these globalization forces.

Albeit improvements in socioeconomic circumstances are themselves favourable, a number of secular trends driven by these changing circumstances have negative effects on health, especially influencing chronic disease risk.

In this thesis, I will examine several factors displaying secular trends all of which have been shown to influence, among other health outcomes, breast cancer risk which is used as a case study.

The thesis is an exploratory investigation of the hypothesis that DNA methylation is one way through which these "exposures" can become biologically embedded and influence disease risk, and especially breast cancer risk.

This hypothesis is mostly driven by evidence that development is plastic so as to allow an organism to respond to its surrounding environment. At certain critical windows of development plasticity is heightened [1,2], but exposures and experiences can leave their mark on the biology of an organism throughout life [3]. Even though modifications in response to environmental signals are vital for proper development, some signals may lead to maladaptive responses [2,4]. These maladaptive responses may not lead to adverse phenotypic changes at the time of exposure, but subtle maladaptive responses may accumulate and persist for years, and they may also interact with subsequent events and/or exposures thus increasing the risk of dysfunction and disease even years after the period of exposure [1,2,4,5].

1.1 Globalization and Secular Trends

Globalization is the process of international integration of cultural aspects. Globalization has become almost synonymous with the term Westernisation which describes the assimilation of Western (mostly North American and Western Europe) traditions and habits by low-middle income nations[6]. Globalization has been the driving engine behind changes in socioeconomic circumstances, behaviours and biology, which have dramatically shifted – to varying degrees - over the past century [7].

These shifts are often termed as *secular trends*. The term *secular trend* has traditionally been used to describe the increasing body dimensions and the associated decrease in the age of onset of puberty observed in succeeding generations [8,9]. Despite the traditional reference to body size and age at menarche with the term secular trends, physical activity and fitness, as well as nutrition and socioeconomic position are all factors that are now considered to have exhibited secular trends [9].

However, given that the direction of change regarding these factors is not always consistent, especially when investigating different populations with widespread geographical and socio-economic backgrounds, *secular changes* is a preferred term to secular trends [8,9], even though both terms are often used inter-changeably.

Here the trends and changes observed in age at menarche and menopause, height, weight, diet and socioeconomic status are reviewed, especially in response to international integration patterns.

1.1.1 Socioeconomic Status

Socioeconomic status (SES) is a term collectively used to describe "measures of resources such as income, wealth, and educational credentials, and the access to goods, services, and knowledge that these resources afford those who have them" [10]. Social status and economic circumstances can refer to individuals, households and groups and can vary across time frames and ethnic groups. There are various measures used to estimate SES, but the ones most often used include highest educational level, income, poverty status, employment and occupational status[10]. With urbanization and globalization trends, as well as with economy shifts, socioeconomic circumstances display their own secular trend, not only over time, but also between and within countries [11].

In the United States, the percentage of children and adults below poverty level has been rising during the past decade, despite a brief decrease observed the decade before last [10] (Figure 1). Similarly, in the European Union (EU), unemployment rates have been on the rise in the past few years, mainly in response to the mounting economic crisis [12] (Figure 2).

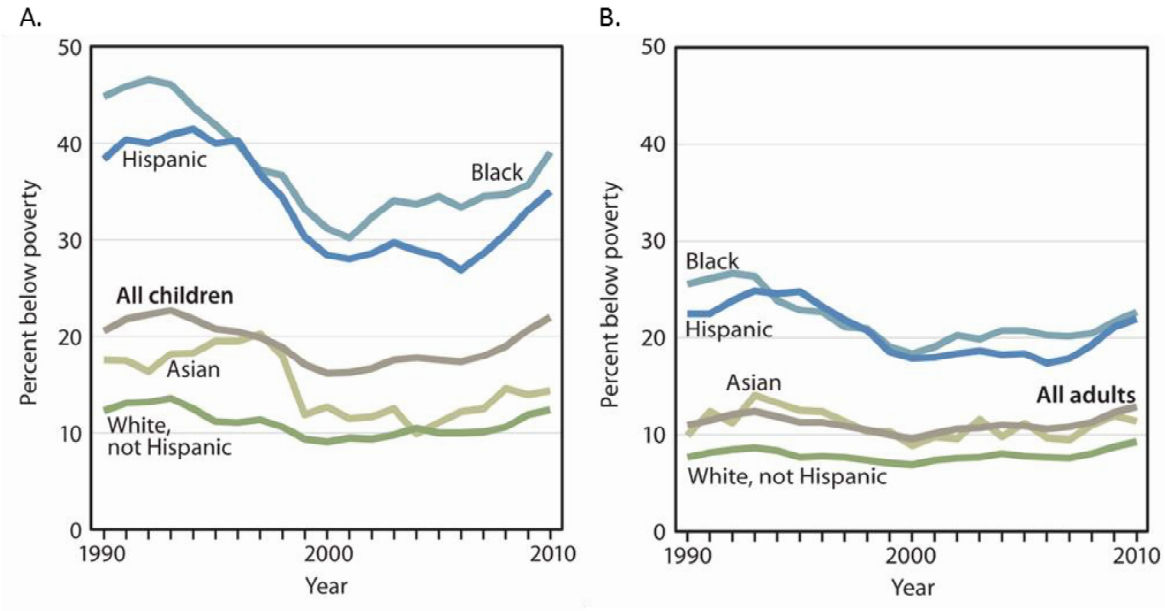


Figure 1 –Trends in Socioeconomic Status.

A) Children and B) Adults 18 years of age and over; by percent of poverty level and race and Hispanic origin: United States, 1990–2010. Reproduced with permission from [10].

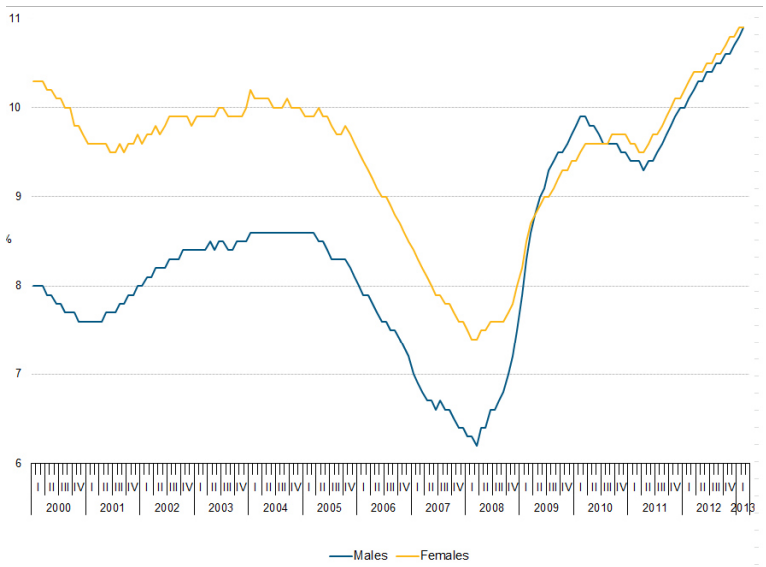


Figure 2 – Unemployment rates by gender, EU, seasonally adjusted, January 2000 - February 2013.

Source: Eurostat. Reproduced with permission from [12].

In addition to the economic crisis, part of the trend in worsening socioeconomic circumstances in high income countries can be explained by urbanisation as it has been linked to increasing poverty levels in urban areas due to internal migration of the high rural unemployment [13]. Additionally, internal migration is associated with over-crowding further impacting on SES [13].

Despite the recent trend towards worsening SES in high income regions such as the EU and the US, developing nations have seen tremendous improvement in socioeconomic circumstances, including living and working conditions, private consumption, social welfare expenditure, access to health services, food production and nutrition, and school enrolment during the past half a century [14].

1.1.2 Nutrition and Dietary Habits

Nutritional habits display their own secular trend towards increased consumption of highly processed foods and drinks as well as alcohol[15] . The widespread increase observed in the consumption of unhealthy food and drink substances has been described as an "industrial epidemic", a term used to describe the potential health effects associated with the commercialization of a product[16].

The secular trends observed in unhealthy food and alcohol consumption can be called an industrial epidemic because they contribute to the burden of non-communicable diseases (NCDs). The popularity of highly processed products has paralleled and contributed to diet-related chronic diseases such as obesity, diabetes, cardiovascular disease and even cancer [17]. In fact, according to the World Health Organisation (WHO) 2010 report on non-communicable diseases, the majority of NCDs examined are strongly and causally associated with tobacco use, physical inactivity, unhealthy diet and alcohol use[17]. In addition, a recent report combining data from 75 countries has reported that soft drink consumption increased globally from 9.5 gallons per person per year in 1997 to 11.4 gallons in 2010, and this increase in consumption was statistically significantly associated with increased risk of overweight, obesity and diabetes [18].

Reports show that the increase in consumption of highly processed foods and drinks, as well as alcohol has reached a plateau in high income countries, however, soft drink and processed food sales continue to rise sharply in low and middle income countries[15]. In fact, richer nations, with high Gross National Product (GNP), who used to take in a large proportion of energy from fats, now seem to turn back to healthier diets, in contrast to low-middle income nations who now consume more animal products and fats than ever before [13]. A marked example is China, where, as incomes rose, carbohydrate diets were abandoned for diets with a high proportion of energy coming from fat [13].

Globalization and the industrial epidemic previously described, spread and make more accessible cheap fat-rich and highly processed foods to a wider market especially in low-middle income countries. In the past few years the production, distribution, and marketing of food have undergone tremendous developments towards a more industrialized process thus being able to cover a very large share of the market, including low-middle income countries [7].

This industrial epidemic has led to the rapid increase of alcohol sales and consumption in low-income and middle-income countries [15]. Processed food sales on the rise in middle and low income countries include vegetable fats and foods high in sugar. In fact, in these countries, there has been a shift from a diet high in fibre and other complex carbohydrates to a diet high in fats and sugars, a shift termed as the "nutrition transition" [13]. The primary cause of this shift is the increased and cheap availability of vegetable fats and sugar-rich foods which in turn has led to an uncoupling of the previously observed relationship between

high income and high intakes of fat [13]. In addition to availability, human preferences towards a more diverse diet are also implicated in the nutrition transition [13].

Despite the recent independency of fat intake from income, the nutrition transition is not independent of urbanization and globalization trends, as urban poorer households consume greater proportions of energy from animal protein, fats, and processed foods and lower proportions from cereal and tubers [19]. Thus, increased urbanization facilitates the nutrition transition.

1.1.3 Weight, BMI and Obesity

With nutrition habits shifting towards highly processed foods, rich in fat and sugars, it is not surprising that the nutrition transition is paralleled by a secular trend in weight, towards increasing prevalence of overweight and obesity, both in children and adults. The secular trend towards increasing weight is a combination of the upward trend in height (described later) as well as an increasing trend in adiposity [8]. Therefore, to delineate the separate effect of height on weight and to concentrate on adiposity, investigations of the secular trend in weight often utilize body mass index (BMI), measured as weight (in kilograms) divided by height in meters squared (weight(kgs)/height(m)^2). Adults are considered overweight when their BMI is above 25kg/m^2 and obese when their BMI is above 30kg/m^2 [20]. Varying definitions have been used to define childhood overweight and obesity [21]. Other adiposity measures often used instead of BMI include waist circumference and skinfold thickness. Despite the tool or definition used, the trend is undeniably one of increasing adiposity, in high and low-middle income world regions as well as in adults and children.

Adult obesity, has been steadily rising during the past 50 years both in the US [22–25] and in Europe[24–26]. In an investigation by von Ruesten et al. (2011) among countries participating in the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort, a dramatic increase in obesity prevalence was observed, even after a short follow-up period of 6 years (Figure 3).

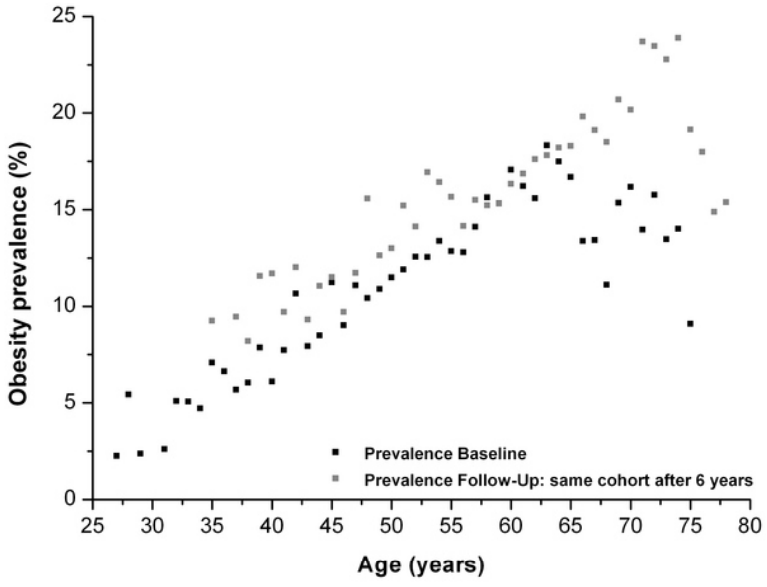


Figure 3 – Association of obesity prevalence and age among women of the Diogenes cohort.

BMI values of individuals of the total Diogenes cohort used for presenting obesity prevalence depending on the age (only prevalence estimates with at least 100 observations per age were considered for the graph). Black squares indicate beaseline prevalence for each age group whereas grey squares indicate obesity prevalence at follow up (6 years). Grey points are

consistently higher than black points indicating a consistent increase in the prevalence of obesity after 6 years for the majority of age groups. Reproduced with permission from [26].

According to available statistics, there appears to be no levelling off or decrease in the rate of change of weight in Europe. Predictions for 2015 indicate that one third of the European population will be overweight assuming a linear increase [26]. In fact, most investigations report an acceleration of the weight gain epidemic especially since 2000 [25].

Figure 4, from Stevens et al. (2012)[25], shows how overweight prevalences have increased in high income countries among females. The highest increase in prevalence is reported for the United States. Figure 4 as well as WHO statistics indicate that the prevalence of overweight and obesity in adults increased with income level, with the prevalence being more than double in high income and upper middle income countries compared to low and lower middle income countries[27].

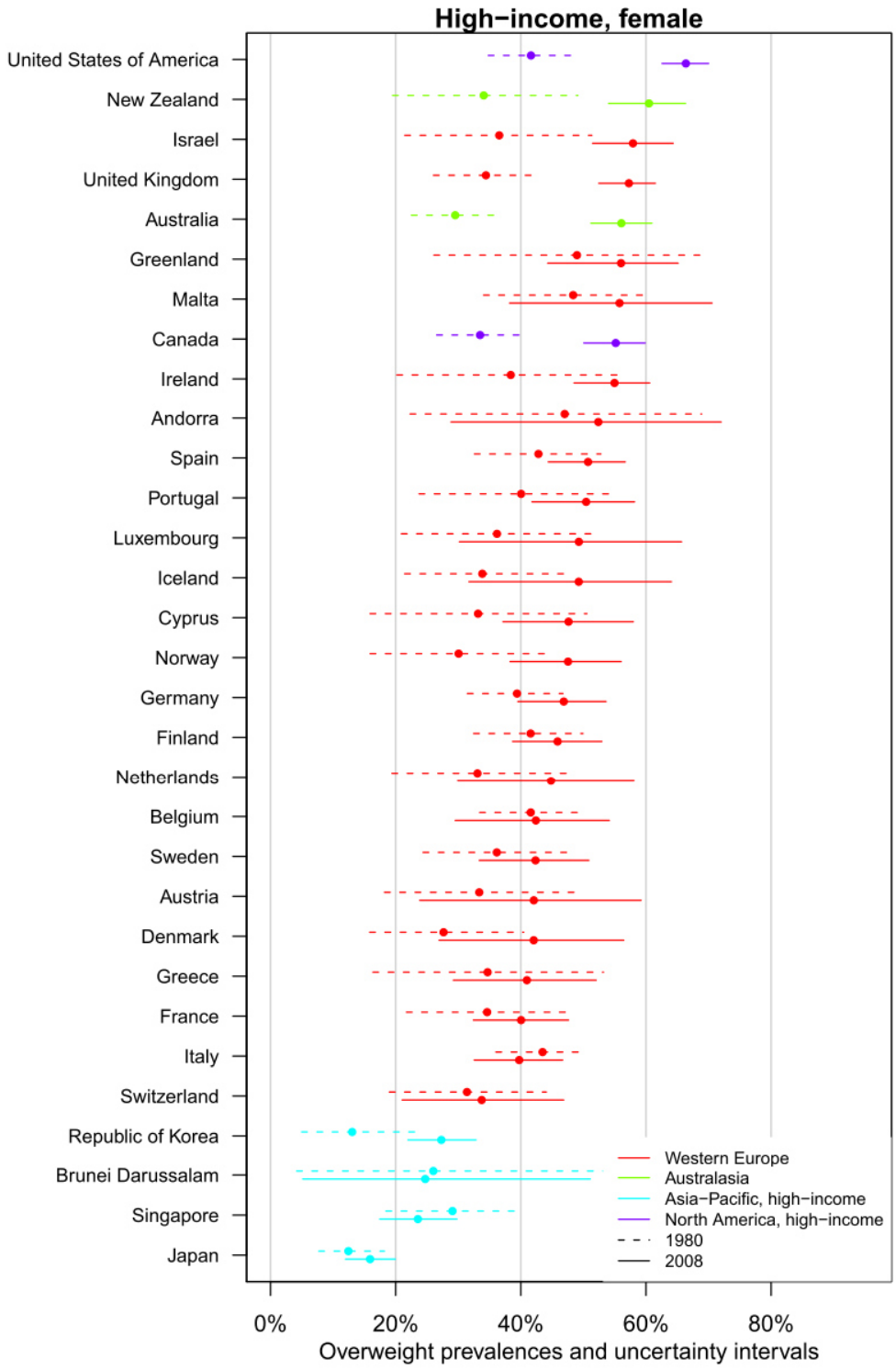


Figure 4 – Prevalence of female overweight and obesity, adults ≥ 20 years, by country and year, 1980–2008.

Reproduced with permission from [25].

In children there is also an evident trend towards increasing weight but the increase has been observed only in the past 30 years [8,28,29]. This increase though has been extensive with almost all countries investigated displaying this trend [30].

In the United States, from 1998 through 2003, the prevalence of obesity increased from 13.05% to 15.21%, and the prevalence of extreme obesity increased from 1.75% to 2.22% [31] especially in children from low income families (Figure 5) . Similarly, rapid increases in the prevalence of overweight schoolchildren are being seen in all EU countries for which data are available[32] (Figure 6).

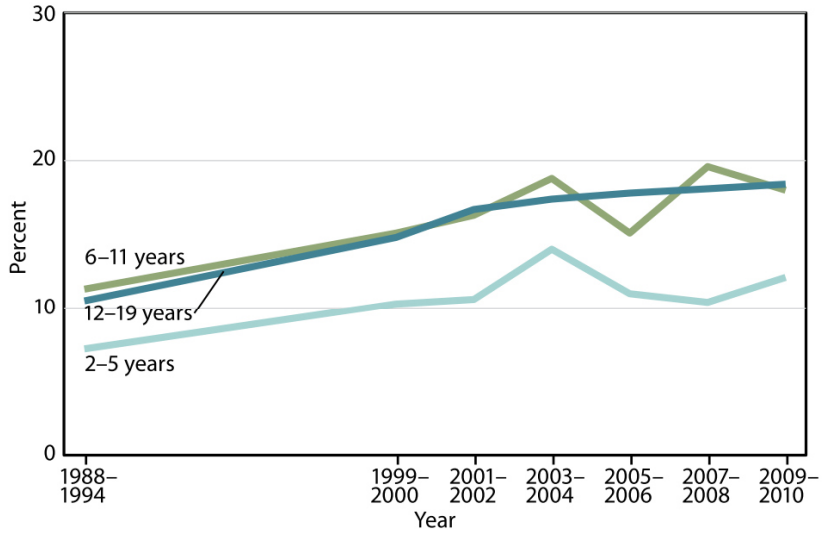


Figure 5 – Obesity among US children by age, 1988-1994 through to 2009-2010.

Obesity defined as BMI at or above the sex- and age- specific 95th percentile BMI cut off points from the 2000 CDC growth charts. Reproduced with permission from [10].

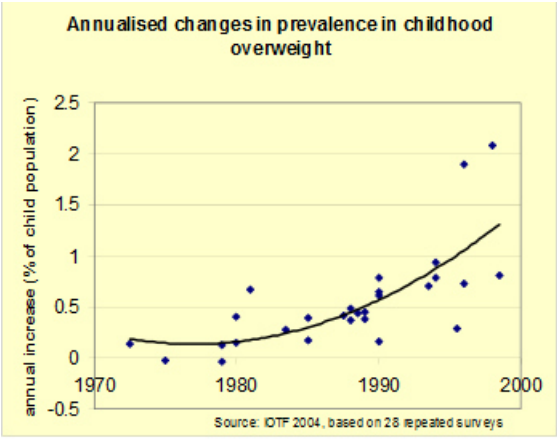


Figure 6 – Changes in prevalence of overweight and obesity per year.

Data are taken from 28 pairs of studies which surveyed comparable populations of children, with at least three years between surveys. Definitions of overweight and obesity were consistent within each pair of surveys but not necessarily between different pairs of surveys. The average annual change in prevalence is plotted against the mid-year between the first and second survey of each pair. The average annual rise in prevalence was under 0.2% of the child population during the 1970s, rose to around 0.4% during the 1980s, rose above 0.6% in the 1990s and is expected to be around 1.0% during the early 2000s. Reproduced with permission from [32].

1.1.4 Age at Menarche / Pubertal Development

Puberty refers to the period of sexual maturation when children attain adult secondary sexual characteristics and reproductive capacity [33]. Pubertal development is characterised by several stages beginning with appearance of breast buds in girls and genital changes in boys, as described by Marshall and Tanner (1969, 1970)[34,35].

Because of the difficulty in ascertaining stage of pubertal development as described by Marshall and Tanner, age at menarche is most often used as a marker of puberty, particularly due to the ease of ascertainment and the reduced likelihood of observational errors [36]. However, it should be noted that age at menarche occurs relatively late along the stages of pubertal development.

In European countries, age at menarche has been steadily dropping from around 16-17 years in the mid-19th century, at a rate of 3.6 months per decade [36]. However, in the last 50 years the trend observed has not been regular, with varying changes in age at menarche in different European Countries. In Eastern European countries the sharply decreasing trend is still observed, however, in several North-Western counties, the trend has slowed down to a non-statistically significant decrease of 1 month per decade. In countries such as Norway, Belgium and Italy some studies suggest that the trend has stopped while in Scandinavian countries [37,38] some studies show that the trend seems to have reversed. However, other recent reports, such as one from the large EPIC study, with data on women aged 35-70 years, report continuing variable decreasing trends in all European countries investigated [39] (Figure 7).

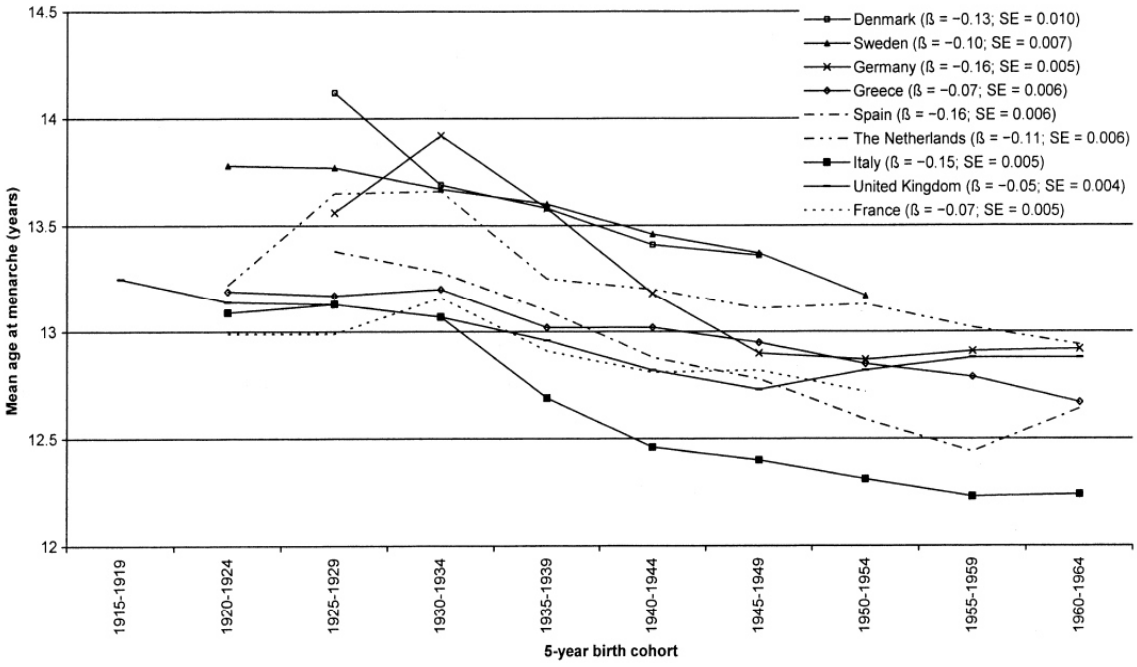


Figure 7 – Mean age at menarche of women born between 1915 and 1964.

Adjusted for socioeconomic status, for 5-year birth cohorts for each participating country (Denmark, Sweden, Germany, Greece, Spain, the Netherlands, Italy, United Kingdom, and France). b is the country-specific regression coefficient from linear regression models, in which the independent variable is 5-year birth cohort and the dependent variable is age at menarche. P-trend < 0.0001 for all countries. SE, standard error. Reproduced with permission from [39].

In the United States, secular advancement in the timing of pubertal development was also observed since the mid 19th century, with rates decreasing as sharply as in European countries [40]. However, during the past 50 years this secular trend has also demonstrated heterogeneity. Using breast development as a marker, the advancement of sexual maturation is only evident in African-American and Mexican-American girls yet less dramatic than in the past 100 years. Mean age at menarche is shown to continue to decrease in all ethnic groups, yet at a faster rate (5.5 months in a decade) in African-American girls compared to white girls (3 months per decade) [41].

The decreasing menarcheal age trend, as well as the variability in this trend observed between countries, can be explained by the multiple factors that act synergistically to determine age at menarche.

These factors firstly include health and nutrition, as reflected through socioeconomic circumstances [9]. For example, better health, nutrition, and sanitation favour earlier sexual maturation and it has been shown that rates of decrease in age at menarche have been more steep in recently developing countries and lower in children from higher socio-economic classes [9]. Therefore trends in SES accompanying the progressive development of low and middle income countries could be one factor driving the age at menarche trend.

In addition, the secular trend in overweight and obesity in children could explain the trend in age at menarche, since the link between childhood obesity and the timing of puberty has been well established[33]. Several mechanistic pathways have been proposed to explain the link between overweight and age at menarche.

Frisch and Revelle first proposed the idea of critical or "permissive" fat mass for the initiation of puberty in 1970 [42]. Despite initial controversy, their hypothesis is now becoming more popular with the discovery of biological pathways that could mediate and substantiate this idea of critical mass. Key in this process is the role of leptin which is secreted by adipocytes and signals to the hypothalamus to trigger gonadotropin secretion and the onset of puberty (reviewed in [33,43]). In fact, Matkovic et al.(1997)[44] have shown an inverse association between serum leptin and age at menarche, with every increase of 1ng/ml in leptin resulting in 1 month earlier age at menarche.

However, despite the "permissive" role thought to be played by leptin, the gradual increase in leptin levels rather than an abrupt increase, suggests that leptin is not the crucial element in determining the time of puberty. Instead, regulators of gonadotropin-releasing hormone secretion might act as the trigger [33].

In another scenario, early life exposures – such as birth weight or rapid weight gain at infancy or even childhood obesity – might act as the trigger to determine onset of puberty. These exposures might affect the hormonal milieu and precipitate in early onset of puberty. Given the crucial interplay of insulin and sex hormone binding protein, it is likely that disturbances in insulin levels affect the levels of circulating hormones and thus pubertal development (reviewed in [33,43]).

In addition, aromatase activity in adipocytes can be elevated in overweight individuals leading to increased conversion of androstenedione to oestrone and testosterone to oestradiol, thus accelerating pubertal development (reviewed in [33,43]).

Despite the biological plausibility of the association, given the majority of the existing literature, it is difficult to ascertain whether the link between age at menarche and obesity is one of causality or co-occurrence. Some studies suggest a link between age at menarche and BMI or weight at puberty [45,46] and others have shown that menarche coincides or is often succeeded by a rapid increase in weight [47]. However, still other studies have investigated the temporality of this association and have shown that onset of menarche is associated with pre-pubertal BMI, supporting causality [48–51]. More detailed study designs are necessary to truthfully delineate the nature of this association. These studies would require a long term perspective and accurate methodologies.

Another factor likely to be implicated in the age at menarche trend is genetic differences. Genetic determinants have a variable distribution showing a North-to-South gradient within Europe[52], and hence they could explain the variability in menarcheal age trends.

Lastly, exogenous exposures such as endocrine disrupting chemicals could also upset hormone balance leading to the earlier onset of puberty. The developmental period of exposure determines the biological effects of these exposures, and exposures at developmentally critical windows, such as infancy and childhood, might have serious potential effects [43]. Whichever the factors impacting on age at menarche, the trend towards earlier sexual maturation in the past century and a half cannot be denied.

1.1.5 Age at Menopause

In contrast to the attention paid to the changes in age at menarche, little attention has been given to the changes in age at menopause. The limited number of studies available provides controversial evidence as to whether there is a secular trend towards increasing menopausal age. Despite several reports showing a constant age at menopause at 50 years of age (reviewed in [53]) more recent reports have shown a significant increase in mean age of natural menopause at the rate of one month per birth year in Swedish women [54]. The secular trend towards later age at menopause was observed also in Italian women [9]. Similar results were also reported for American women [55]. The latter study also reported a steady increase in mean reproductive period from 36.1 years to 37.7 years in women born 20 years apart [55].

1.1.6 Height

Since the mid 19th century, male conscript data show a visible expression of the secular trend in adult height with grown-up children being taller than their same sex parents [8,9]. The secular trend towards increasing adult height has been evident in several countries worldwide, but, similar to the menarcheal age trend, the rate of change has slowed down significantly in Northern European countries, whereas the upward trend continues in Southern European countries (reviewed in [8,9]) (Figure 8).

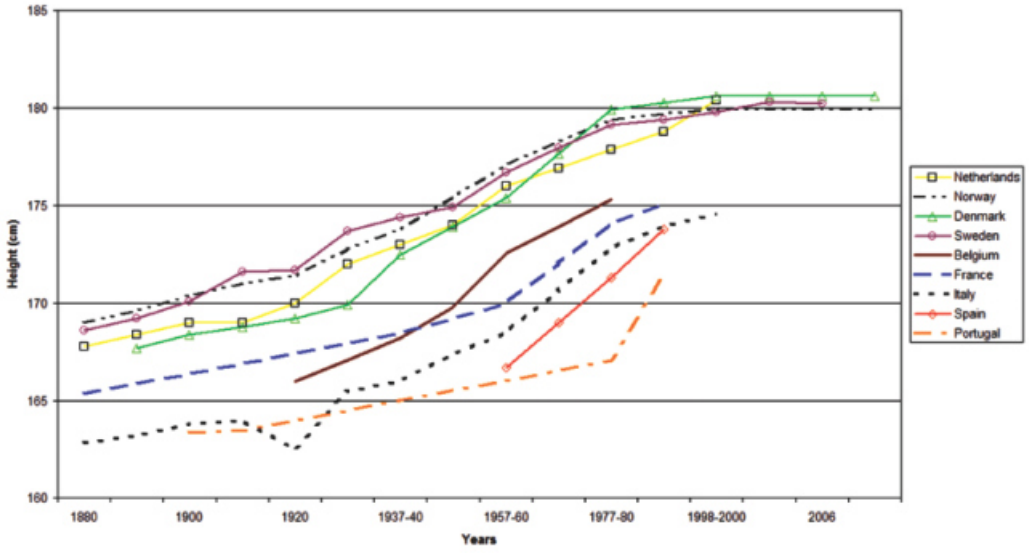


Figure 8 – Average height (cm) at time for conscription per country from 1880 to 2006.
Data from: Chamla, 1964, 1983; Lindgren, 1998; Rebato, 1998; Padez, 2003, 2007; de Beer, 2004; Komlos & Baur, 2004; Statistics Norway, 2006; Werner, 2007; Statistics Denmark, 2007; Statistics Netherlands, 2007. Reproduced with permission from [9].

The height trend in children is similar to that observed in adults, but, not surprisingly, it is much more dependent on child's age [8]. More specifically, the rate of change seems to be larger around puberty [56]. Surprisingly, there is no observable trend in birth length suggesting that environmental factors may impact on growth after birth.

Of course, one cannot deny the strong genetic influences on height determination but the rapid increase in height especially over the span of only half a century is too steep to be explained by genetic factors alone. Instead, it has become obvious that adult height is dependent on a combination of genetic and environmental factors that interact to influence growth. Environmental factors influence height indirectly through biological factors such as infection, nutrition and stress, therefore improved nutrition and health could explain the trend towards increasing height [57]. The variability in socioeconomic circumstances between European countries could explain the variations of this secular trend observed. For example, in Northern European countries, where socioeconomic conditions are on average better than in Southern Europe, the peak of the rate of change in height has passed and is now beginning to level off. On the

contrary, in Southern European countries, height continues to increase as socioeconomic conditions continue to improve [9]. Reports that European countries have a higher biological standard of living, could also explain why the secular trend in the US has not kept up with the increase in height in European populations [9].

1.2 Secular Changes and Health Effects

The secular trends described above are not without serious health consequences.

1.2.1 Socioeconomic Status Health Effects

An undeniable social gradient in health exists depending on the SES of individuals, groups and countries. More specifically, health is progressively better the higher the socioeconomic circumstances [11,58]. Adler et al. (1994, 1999) have demonstrated a positive linear relationship between decreasing socioeconomic position and mortality as well as morbidity [58,59] (Figure 9). In the same publication, an even steeper linear relationship was observed between socioeconomic position and prevalence of chronic diseases, with prevalence decreasing at each higher socioeconomic position [58,59]. The strongest and most consistent relationships observed were for cardiovascular disease, metabolic syndrome and diabetes, arthritis, tuberculosis chronic respiratory disease, gastrointestinal disease, and adverse birth outcomes (summarized in [58]). For cancer the strength and direction of the association will depend on the cancer type examined and whether one is investigating incidence or mortality. For example, for breast cancer, studies show that breast cancer incidence is increased as SES increases, but breast cancer mortality is decreased [58].

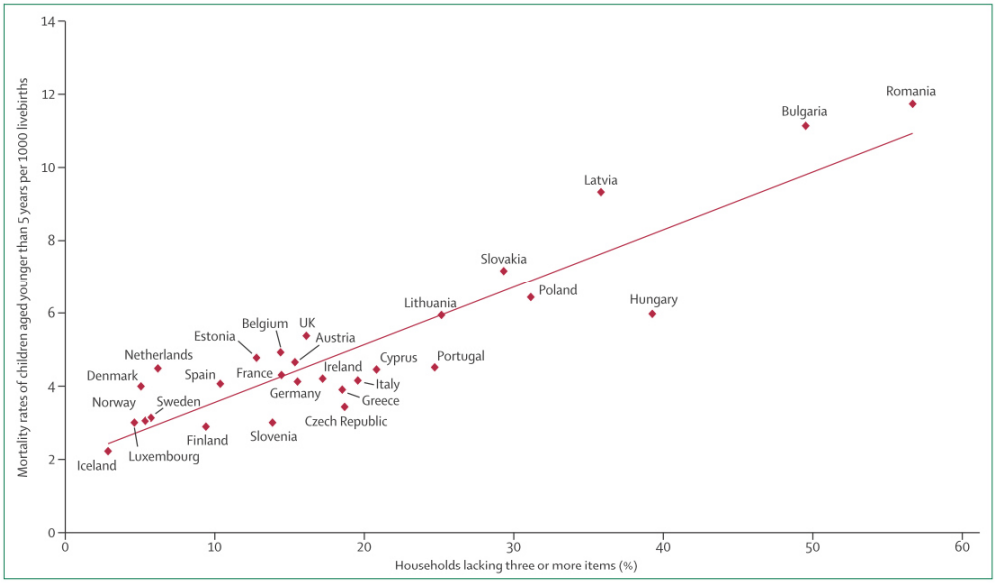


Figure 9 – Mortality among children younger than five years of age and percentage of deprived households (lacking three or more essential items) in selected countries in the WHO European Region.

Reproduced with permission from [11].

In addition to chronic disease morbidity, chronic disease risk factors, such as smoking, cholesterol and physical inactivity, also show significant negative relationship with SES.

The mechanisms through which socioeconomic circumstances influence health have been the subject of much research during the past few years. SES can influence disease risk directly either through exposure to carcinogens and pathogens or through central nervous system and endocrine responses to psychological influences[58]. However, SES can also indirectly impact on health through performance of health-relevant behaviours. Recent research has evidenced a strong attenuation of the relationship between SES and mortality, when accounting for health behaviours such as smoking, diet, alcohol consumption, and physical activity [60,61]. The mechanisms through which SES can influence health are summarized in Figure 10.

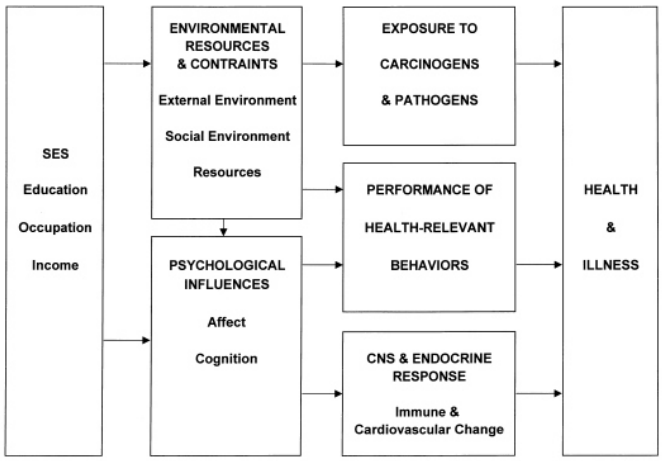


Figure 10 – Model of the pathways through which SES influences health.

Reproduced with permission from [58].

In addition to adult SES, early life social conditions, including prenatal socioeconomic status, are also associated with adult chronic disease [62]. Low prenatal SES has been associated with stroke incidence [63], and with cardiovascular disease (CVD) mortality [64] in adult life. This supports the idea that socioeconomic circumstances become biologically embedded or "get under the skin" through a mechanism that can influence health even many years after exposure.

1.2.2 Nutritional Health Effects

Changing nutritional patterns, and more specifically the shift towards a diet rich in highly processed fatty and sugary foods has had several negative health outcomes.

Firstly, the nutrition transition has contributed, among other factors, to the secular trend towards increasing overweight and the epidemic of obesity. Overweight/obesity, is one of the major attributable causes of death, being responsible for more than 2.3 million deaths worldwide, yearly [65] and the morbidity implications of overweight and obesity are discussed in the next section.

Dietary habits rich in highly processed, salty and fatty foods are also responsible for the rising prevalences of hypertension and hypercholesterolemia which are themselves responsible for 7.6 and 3.9 million deaths annually, respectively [65]. The high mortality rates arise primarily from heart problems associated with these conditions. Hypertensive heart disease is a term used to describe heart failure and other heart complications when hypertension is the underlying cause of heart disease [66]. Hypertension can result in progressive myocardial hypertrophy, greater collagen volume fraction in the myocardium, microvascular disease and endothelial dysfunction as well as increased arterial stiffness all of which can contribute to heart failure [67,68]. Furthermore, hypertensive heart disease associated atrial fibrillation and ventricular arrhythmias increase one's risk of atherosclerosis, ischaemic heart disease and myocardial infarction [68]. Hypercholesterolemia and dyslipidaemia are also associated with the build-up of atherosclerotic plaques and coronary heart disease (CHD), and dietary management with a low-fat low-cholesterol diet is essential for the management of the disease [69,70].

The changing nutritional patterns towards a more westernized type of diet, have also been coupled with low fruit and vegetable intake [13]. Close to 2.6million deaths annually are attributed to diets with inadequate fruit and vegetable intake [65]. This is because inadequate intakes of fruit and vegetables are associated with increased risks for ischaemic heart disease, cerebrovascular disease, lung cancer, stomach cancer, colon and rectum cancers, and oesophageal cancer [71]. The findings on breast cancer are more controversial but results from some large prospective cohort studies and meta-analyses suggest that high fruit and vegetable intake reduces breast cancer risk [72–76], with more pronounced effects on oestrogen receptor negative cancers. This effect may be explained by low folate intakes, since inadequate folate intake has been found to be associated with breast cancer risk [77–82]. The protective effects of folate are more pronounced in women who consume moderate amounts of alcohol [72,83–85] and low folate intake is more detrimental when alcohol consumption is above 4 gms per day [86]. Furthermore, it was recently shown that increased caloric intake, often associated with intakes of highly processed, fatty foods, was associated with 60-70% increased risk of breast cancer in pre-menopausal women [87].

A diet rich in processed sugars is also associated with type II diabetes, a condition directly responsible for around 1 million deaths worldwide annually, and a risk factor for other serious life-threatening conditions [65]. Diabetes type II is often preceded by insulin resistance syndrome, also called metabolic syndrome. Insulin resistance syndrome and type II diabetes are associated with increased CHD [88] as well as with various cancers such as liver, pancreatic, colorectal, endometrial and postmenopausal breast cancers [89].

Lastly, increased alcohol consumption is one of the world's leading health risks [90]. More than 2 million deaths worldwide in 2004 and about 4.5% of the global burden of disease are attributed to increased alcohol consumption. These diseases include neuropsychiatric disorders such as epilepsy, gastrointestinal diseases such as liver cirrhosis and pancreatitis, increased risk for cancers of the colorectum, female breast, larynx, liver, oesophagus, oral cavity and pharynx, as well as cardiovascular diseases such as hypertension [90].

An example of a well-balanced diet, rich in complex carbohydrates and with moderate consumption of animal fats and alcohol is the Mediterranean diet. The findings of a few large prospective cohort studies support that the Mediterranean diet is protective for breast cancer [91,92].

In addition to large prospective studies, several case-control studies have evidenced a protective effect of the Mediterranean diet on breast cancer risk. An investigation in the Cypriot population, a purely Mediterranean population, has revealed that a Principal Component Analysis (PCA) derived dietary pattern, rich in foods at the bottom of the Mediterranean Diet pyramid such as fruits, vegetables, fish and legumes, decreases breast cancer risk[93]. The results of this investigation, which form a part of the work carried out for this thesis, have been published and the paper is attached (Appendix A).

The evidence on the protective role of the Mediterranean diet and its components, as well as possible mechanisms conferring protection, are reviewed in a book chapter titled "The Mediterranean diet and breast cancer risk", written for publication in "The Mediterranean Diet: An Evidence-Based Approach To Disease Prevention". This is a new book which will be published by Academic Press, an imprint of Elsevier, and will be co-edited by Prof. Victor R. Preedy, Kings College London. A copy of the chapter, that was a product of the work carried out for this thesis, has been accepted for publication and is attached (Appendix B).

1.2.3 Overweight and Obesity Health Effects

As mentioned previously, overweight/obesity, is one of the major attributable causes of death worldwide [65]. Obesity and overweight are associated with increased risk of developing insulin resistance syndrome and diabetes type II with the above mentioned serious adverse health outcomes [24]. Overweight and obesity are also associated with traditional risk factors for CHD such as hypertension and dyslipidaemia whose health effects are previously discussed [94].

In addition to the above, several cancers are partly attributed to overweight and obesity. The available evidence is convincing regarding associations between overweight/obesity and adenocarcinoma of the oesophagus and the gastric cardia, colorectal cancer, postmenopausal breast cancer, endometrial cancer and kidney renal-cell cancer [89]. The evidence for cancers such as liver, gallbladder, pancreas, thyroid gland and in lymphoid and haematopoietic tissue cancers is less consistent and thus less convincing[89].

Overweight/obesity also increase risk for musculoskeletal disorders, especially osteoarthritis, rheumatoid arthritis, spondyloarthropathy, and fibromyalgia [95].

Children who are overweight or obese also face increased risks for developing insulin resistance syndrome and type II diabetes, cancer, heart conditions and neuromuscular disorders in adult life [20,96].

It is important to state here that the health effects described for nutrition and overweight/obesity are patterns derived from epidemiological studies, but the causal links between diet, obesity, physical exercise, insulin resistance, metabolic syndrome, and non-communicable diseases are far from being understood. The mechanisms linking these exposures and biological phenotypes have yet to be delineated thus it is difficult to deduce which comes first and what causes which. Therefore, the evidence regarding dietary and overweight/obesity health effects needs to be interpreted with caution, bearing in mind the interplay between the extrinsic and intrinsic components of such patterns and pathways.

1.2.4 Age at Menarche and Age at Menopause Health Effects

Earlier age at menarche, as well as later age at menopause, are two of the most well established risk factors for breast cancer [97–99]. The mechanisms through which an earlier age at menarche and a later age at menopause increase breast cancer risk are not completely understood but greater cumulative endogenous oestrogen exposure because of a longer reproductive period could be affecting breast cancer risk [100]. This idea is supported by a recent meta-analysis of studies which shows a greater protective effect of a later age at menarche for ER+/PR+ cancers compared to ER-/PR- cancers [98]. However, much research is still needed to delineate the mechanisms through which age at menarche impacts on breast cancer risk.

In addition to breast cancer, an earlier age at menarche has been associated with increased risks for adult obesity [101], diabetes [102], cardiovascular risk factors [103], altered psychosocial behaviours [104] and mental health problems [36].

1.2.5 Height Health Effects

Perhaps the secular trend least associated with adverse health effects is increasing height. Despite evidence that taller women are at increased breast cancer risk[105], a taller height – at childhood or adulthood – is associated with better health and greater cognitive ability throughout life [106]. However, these improved health outcomes may reflect better nutrition and health during early life which lead to taller children and predispose to a more favourable adult health [106].

1.2.6 Secular Changes and Breast Cancer

One of the adverse health outcomes common to the secular trends described above is cancer, and more specifically breast cancer. Yet, the mechanisms through which these factors influence breast cancer risk are not completely understood.

1.2.6.1 Adult SES

Socioeconomic status in adulthood is positively associated with breast cancer, in that a higher SES is associated with increased breast cancer risk [107]. Adult SES may impact on breast cancer indirectly by determining reproductive and lifestyle choices which in turn influence risk. In fact, SES is associated with a number of breast cancer risk factors such as age, parity, age at first birth, BMI, height, alcohol consumption, age at menarche and family history of breast cancer [107], and adjusting for these factors in a model investigating the relationship between SES and breast cancer can attenuate the association to the point of non-significance [107]. However, it is also possible that the increased incidence in individuals and countries with higher SES is related to better access to healthcare and thus better and earlier diagnosis, and not to a biological effect.

Similarly, early life socioeconomic circumstances can influence breast cancer risk factors, which in turn increase or decrease breast cancer risk. For example, better health and nutrition during early life are associated with a taller height and an earlier onset of menarche which increase breast cancer risk [9,105,106]. Also, unfavourable socioeconomic conditions and poorer education are associated with increased parity [108], which decreases breast cancer risk [98].

1.2.6.2 Nutrition

The nutrition transition, including excessive intakes of fat and sugar as well as inadequate intakes of beneficial fruits and vegetables, could also influence breast cancer risk, directly or indirectly.

Folate, found largely in fruits and vegetables, is a key player in the one carbon metabolism pathway. This pathway is not only responsible for the methylation of biological molecules such as DNA and RNA, but it is also essential for DNA synthesis and repair. This is because folate derived one-carbon groups, are needed for the *de novo* synthesis of pyrimidine, thymidylate, as well as the purines [109]. There is evidence that folate deficiency may lead to global hypomethylation [110] which is an event frequently observed in breast tumours, and is associated with activation of oncogenes as well as chromosomal instability through repetitive

element activation [111]. In addition, given folate's role in maintaining DNA integrity, as well as in synthesis, and repair, folate deficiency may compromise the reliability of these processes leading to tolerance in DNA damage which is also associated with carcinogenesis[112].

Fruits and vegetables are also rich in carotenoids, such as retinol, retinyl pulminate, α-carotene, β-carotene, β-cryptoxanthin, lutein and lycopene. These substances have been shown to decrease proliferation and promote cell differentiation of epithelial cells in culture [113,114]. In addition, carotenoids have anti-oxidant function and can protect macromolecules such as DNA from damage from Reactive Oxygen Species (ROS)[112].

In addition to folate and carotenoids, flavonoids are also abundant in a diet rich in fruits and vegetables. Flavonoids, such as catechins, are polyphenolic compounds found in plants, and have been shown to be biological response modifiers. High flavonoid intake has been consistently shown to decrease breast cancer risk, and this protective effect might stem from their effect on ovarian sex hormone synthesis, or their other biological activities such as antioxidant, anti-proliferative, and antiangiogenic activities (summarised in [115]). In addition, flavonoids have been shown to inhibit the effects of cytokines, growth factors, and several other enzymes involved in carcinogenic processes (summarised in [115]).

Phytooestrogens, plant-derived xenooestrogens, are found in large amounts in diets with a high content in fruits, vegetables, and fibre. Phytoestrogens have a similar molecular structure to oestrogens which allows them to have ooestrogenic or anti-oestrogenic effects by binding to oestrogen receptors [116]. Phytoestrogens can also modulate the bioavailability of sex hormones by stimulating or inhibiting the action of enzymes such as the Sex Hormone Binding Globulin (SHBG)[117]. The anti-ooestrogenic effects of phytooestrogens have been shown to confer a protective effect for breast cancer, however the evidence has not been always consistent [118].

On the other hand, excessive fat and sugar intake are probably not directly associated with breast cancer [119,120], but affect breast cancer risk through their association with increased adiposity and diabetes risk.

1.2.6.3 Overweight/obesity

Overweight and obesity can influence carcinogenesis through a variety of mechanisms reviewed in detail elsewhere [121–124]. In brief, one of the key mechanisms involves excess androgen to oestrogen hormone conversion and secretion by hypertrophied adipose tissue[122]. This results in increased bioavailability of estradiol which has proliferative and anti-apoptotic effects on target organs such as the breast and endometrium [125]. Hypertrophied adipose tissue, also releases inflammatory adipocytokines which affect cell growth, survival, proliferation and angiogenesis[121–124]. Where obesity is accompanied by insulin

resistance, insulin directly promotes cell growth but it also indirectly stimulates the release of other hormones which act as growth factors[121–124]. In addition, obesity can influence carcinogenesis through increased circulating leptin levels which have been shown in vitro to stimulate proliferation of neoplastic and pre-neoplastic cells and to trigger the release of inflammatory cytokines [122,126]. Despite the number of possible mechanisms through which obesity influences carcinogenesis and the inevitable interplay between these compex pathways, there is consensus that obesity contributes to increased breast cancer risk and more specifically, post-menopausal breast cancer risk.

1.2.6.4 Age at Menarche

As already discussed, the mechanisms through which an earlier age at menarche and a later age at menopause increase breast cancer risk are not completely understood but greater cumulative endogenous oestrogen exposure because of a longer reproductive period could be affecting breast cancer risk [100]. This idea is supported by a recent meta-analysis of studies that shows a greater protective effect of a later age at menarche for ER+/PR+ cancers compared to ER-/PR- cancers [98]. On the other hand, early life exposure to endocrine disrupting chemicals (EDCs) could affect the hormonal milieu in a way that precipitates both to an earlier pubertal development and, independently, to an increased breast cancer risk[43]. For example, animal models have evidenced that early life exposure to an EDC, Bisphenol A, which binds to estrogen receptors, elevates testosterone as well as estradiol and affects neuroendocrine signalling, can accelerate and alter mammary gland development and increase breast cancer risk [127–129]. In the same way, other EDCs can accelerate pubertal development leading to an earlier age at menarche, and, independently, can also increase breast cancer risk.

1.2.6.5 Height

Lastly, it is possible that height may increase breast cancer risk directly because of the increased number of cells, including stem cells, in taller people and hence increased likelihood of malignant transformation [130]. However, it is more likely that height affects breast cancer risk indirectly, through factors that act during early life to determine both height and breast cancer risk. Height is determined by both genetic and environmental influences. These environmental exposures can include childhood nutrition and infections, as well as exposure to growth factors such as insulin-like growth factors (IGFs), both in childhood and in adult life [130]. These factors influence both height and risk of breast cancer and may explain the association between height and breast cancer risk.

1.2.7 Biological Embedding of Trends and Exposures

For early life trends and exposures, such as exposure to EDCs, socioeconomic conditions, nutritional patterns, and overall health, to influence biological responses sometimes even years after the period of exposure, they need to "get under the skin" and become biologically embedded in an organism [3,131]. Several mechanisms for biological embedding of early life exposures have been proposed. For example, early life SES can become biologically embedded through several processes such as deregulation of hypothalamic-pituitary-adrenal axis (HPA), inflammatory processes, and neural function and structure [3,131]. Recent reports show that one possible mechanism though which early as well as later life exposures can become biologically embedded and influence such processes is epigenetics, especially when the exposure occurs at times of developmental plasticity.

Development is plastic so as to allow an organism to respond to its surrounding environment, especially very early on in development when cells and tissues are differentiating [4]. At these critical windows of development, plasticity is heightened since proper growth and differentiation depends on nutrient availability, chemical signals from surrounding tissues, as well as environmental signals [1,2]. These critical windows are time periods of rapid change when biological processes are most easily modified by internal and external factors and exposures, and these intrinsic changes are often irreversible [132]. This does not, however, mean that biological embedding of experiences is confined to an early critical period. Studies have shown that exposures and experiences can leave their mark on the biology of an organism throughout life [3]. In fact, certain exposures can have a strong effect on the biology of an organism during a sensitive period, but they can affect biological processes, weakly, even outside sensitive periods [132].

Even though modifications in response to environmental signals are vital for proper development, some signals may lead to maladaptive responses [2,4]. These maladaptive responses may not lead to adverse phenotypic changes at the time of exposure. However, subtle maladaptive responses may accumulate, persist or even be re-inforced for years, increasing the risk of dysfunction and disease even years after the period of exposure [1,2,4,5]. This hypothesis was first termed "The Thrifty Phenotype Hypothesis" [133] and in contrast to "The Thrifty Genotype" hypothesis, accounts for the rapidly increasing incidence of diseases, such as breast cancer [1].

The first evidence for "The Thrifty Phenotype" hypothesis came from a series of reports linking birth weight and early life nutrition to long term disease risk [134–139]. Despite the controversy that followed and the discrepancy of results depending on whether the disease or some surrogate of the disease was the defined outcome, a large body of evidence confirms these findings [140]. To further support this hypothesis, several reports demonstrated that individuals born during the Dutch famine or the Great Chinese Famine tend to

have lower birth weight, increased female BMI, hypertension, and increased plasma glucose and hyper-triglyceridemia compared to unexposed same sex siblings [141–144].

It is important to note that the responses to an environmental signal – whether adaptive or maladaptive – are stage dependent [141] and can persist for generations [145]. These stage dependent, heritable, and persistent phenotypic effects can be explained by epigenetic modifications imposed by environmental signals, which are themselves heritable, time dependent, and tissue as well as sex specific [3,4,146,147]. The epigenome has the capacity to enable an organism to respond to environmental signals due to its control of gene expression without altering the DNA sequence. The epigenetic modifications that impact on gene expression include DNA methylation, histone modifications, miRNA expression, and nucleosome remodelling [148] and just like other systems, they too are especially sensitive during the critical windows of developmental plasticity. Therefore, early and later life social conditions and exposures can influence disease risk even years after discontinuation of the exposure by impacting on the epigenome, making epigenetics a possible mechanism through which the factors undergoing secular change can influence breast cancer risk[3].

1.3 Breast Cancer and Epigenetics: a narrative review

Lending support to the hypothesis that the factors described above can influence breast cancer risk through epigenetic mechanisms is the fact that breast cancer is characterised by an aberrant epigenome.

1.3.1 The Epigenetics of Breast Cancer

It is known that breast cancer initiation and progression is triggered, by accumulation of pathologically altered gene function. These deregulated genes then affect many important cellular networks such as cell cycle control, apoptosis, DNA repair, detoxification, inflammation, cell adhesion, and/or migration, which are considered key in cellular progression from normal to neoplastic [149].

For many years, cancer was considered as "a genetic disorder of fatal acquisition of multiple mutations" in key genes that regulate the important networks outlined above. These mutations can lead to the silencing of genes that keep the cell functioning normally, e.g. tumour suppressor genes such as *TP53* and *BRCA1*, or to the overexpression of genes that drive the cell towards neoplasia e.g. oncogenes such as *MYC*. In either case, the cell is driven towards malignant transformation [149]. This is known as the somatic mutation theory.

However, this theory has been revolutionized since it is now well accepted that several models of carcinogenesis exist[150]. One of these models is based on non-genotoxic mechanisms, with epigenetics as its main features [150]. In fact, in this non-genotoxic model, epigenetic malfunction can contribute to carcinogenesis just as much as genetic alterations. Epigenetics refers to "heritable changes in phenotype or gene expression that are caused by mechanisms other than changes in the DNA sequence" [149]. Epigenetic mechanisms control crucial biological processes such as X chromosome inactivation, genomic imprinting, reprogramming of genomes during differentiation and development, and RNA interference [148]. Epigenetic events such as gene silencing via hypermethylation can be added to genetic events, such as mutations, thus inactivating both alleles of a gene, leading to neoplasia, through Knudson's two hit model [151-152].

As far as transcriptional regulation is concerned, four epigenetic mechanisms have emerged as having the most crucial roles: DNA methylation, histone tail modification, miRNA expression, and nucleosome remodelling. It is important to note that there is a strong interplay between these processes and it is this interplay that defines the 'epigenetic landscape' or 'epigenome' – "the epigenetic status that determines the way a single eukaryotic genome may manifest itself in different cell types and developmental stages, and that, if aberrant gives rise to cancer and other diseases" [153]. In fact, epigenetic changes are involved in every step of tumour development and progression [135-136]. This review focuses on DNA methylation.

In the human genome, 4% of cytosine residues are methylated on the 5th carbon of the pyrimidine ring, post replication, to produce 5-methylcytosine (5mC), and most of these are found in CpG dinucleotides. In this manner, 75% of CpG dinucleotides in mammalian genomes are found to be methylated [111,156]. Most

CpG dinucleotides are found in islands in the promoter regions of approximately 60% of the genes in the mammalian genome [153] or in regions of large repetitive sequences like retro-transposon elements.

A high percentage of CpG dinucleotides in repetitive sequences are methylated to inhibit activation and maintain chromosome stability, but CpG sites in CpG islands associated with gene promoters are usually unmethylated. These unmethylated promoter regions allow for active gene transcription [157-158] and also have a role in cell differentiation. CpG dinucleotides in the promoters of housekeeping genes are generally unmethylated. On the contrary, CpG islands on tissue specific genes are found methylated to inhibit or downregulate gene transcription depending on the type of tissue [159-160].

Recently it was established that DNA methylation involved in differentiation, can occur not only on CpG islands but also on regions surrounding CpG islands, termed as CpG island 'shores'[161–163]. These shores are regions of relatively low CpG density compared to CpG islands, located within gene promoters. In addition to CpG methylation, non-CpG methylation and 5-hydroxymethylcytosine (5hmC) are also part of the DNA methylation pattern.

The pattern of methylation is strictly maintained by the cell, except in cases of altered gene expression in which case the methylation pattern might change temporarily. However, in cancer cells, methylation patterns can be greatly disturbed. In fact, aberrant DNA methylation patterns seem to define specific tumour types [164-165]. To discern whether DNA methylation from a neoplastic cell is disturbed, the gold-standard method would be to compare the tumour DNA methylation to a relatively pure population of cells which is known to be the cell of origin of the tumour[166]. However, since this is usually not possible, the majority of studies investigating cancer associated aberrant methylation, compare tumour tissue to adjacent normal tissue or affected tissue from a case to non-tumour same-tissue from a control. The limitations of these approaches are further enhanced by the lack of a normal, tissue-specific reference [167]. Until the major sequencing initiatives enabled by the epigenome technology revolution help us understand what is "normal" in terms of the epigenome, such methodological caveats are unavoidable [167].

1.3.1.1 Global Genomic Hypomethylation

It has been shown that in several cancers, including breast cancer, whole genome methylation is decreased compared to normal tissue [168]. Whole genome methylation is most commonly assessed using surrogate repetitive elements such as long interspersed nuclear element-1 (LINE-1) and Alu repeats, not all of which are accurate reflections of whole genome methylation [169]. However, the evidence that genomic 5mC content is reduced in the blood of cancer patients compared to the blood of controls or in cancerous tissue compared to normal tissue is convincing [169–172]. How early this occurs might depend on the neoplastic

tissue and on the type of tumour [166]. In cancerous breast tissue, hypomethylation relative to normal tissue has been shown to mainly affect endogenous retro-transposons such as repetitive DNA sequences and pericentromic satellite DNA, regions that are heavily methylated in normal cells. For example, LINEs that are methylated in normal cells, become hypomethylated and activated in breast cancer cells and can reposition themselves elsewhere in the genome, causing insertional mutagenesis, transcriptional interference, and genomic instability[149,166,169]. Also, satellite DNA sites were shown to be frequently hypomethylated in human breast adenomas[166,173,174]. Further to activating repetitive DNA sequences, DNA hypomethylation might also contribute to translocations of these hypomethylated sequences, by loosening chromatin packaging [166,175,176].

To support global hypomethylation's effects on the genome, a study in breast cancer cell lines has correlated chromosomal instability events with the degree of global hypomethylation [177,178]. In many human tumours, areas that lack DNA methylation suffer chromosomal structural changes [174,179]. In addition to transposable element reactivation and chromosomal instability, hypomethylation can also contribute to carcinogenesis through loss of imprinting [148,180,181] . Lastly, hypomethylation in tumour tissue was shown to appear early and precede mutation and deletion events in cancer [182]. Hypomethylation can be associated with tumour progression, as in the case of ovarian cancer, or it can be involved in early tumour development, as in the case of breast cancer.

The cause of global hypomethylation is still debated but it is now understood that the process of methylation is dynamic and complex and DNA methyltransferases (DNMTs) are involved both in the addition and in the removal of methyl groups [183]. There are three types of DNMTs: DNMT1 is involved in maintenance of methylation through DNA replication, DNMT2 is involved mainly in tRNA methyltransferase activity, and DNMT3A and DNMT3B which are mainly expressed in the embryo and are responsible for *de novo* methylation. One proposed mechanism for DNA hypomethylation is passive demethylation when DNMT1 is inhibited and methylation is not maintained through rounds of DNA replication[183]. DNMT1 may be chemically inhibited or its expression might be reduced in cancer cells, due to deregulation of DNMT1 expression by tumour suppressor genes which are themselves deregulated [184].

However, active demethylation is also possible. One hypothesis is that active global hypomethylation is caused not by a decrease in DNMT [178,185] but by a global increase in demethylase activity [186]. Another proposed mechanism is the removal of methyl groups through oxidation or hydrolytic attack [187]. Yet another hypothesis is that enzymatically inactive DNMT splice variants, compete with enzymatically active forms for the same binding site, thus inhibiting DNA methylation[149,188,189]. It is also likely that DNA demethylation involves the DNA repair machinery either through removal of the methylated base and substitution by an unmethylated one, or by removing the entire DNA patch and then filling it in with new nucleotides[183]. Lastly, Tet proteins may also mediate active DNA demethylation by hydrolysing methyl-

cytosines to produce 5-hydroxymethyl-cytosine, blocking DNMTs and inducing chromatin remodelling [183,190,191].

1.3.1.2 Gene Specific Hypermethylation

In addition to genome-wide hypomethylation, in cancer cells, CpG sites that are normally unmethylated in non-cancer cells, may become methylated leading to the silencing of genes. As stated, this silencing of an allele due to hypermethylation can be coupled to the genetic silencing of the other allele, leading to neoplasia through Knudson's two hit model[179,192,193]. In fact, genes that are implicated in familial forms of cancer when mutated in the germ line, have been shown to be hypermethylated and silenced in sporadic cancers [179]. For example, mutations in the *BRCA1* gene are of major importance in the aetiology of familial breast cancer. However, it has also been shown that in the tumours of women with sporadic breast cancer, *BRCA1* gene is often hypermethylated[194,195], and the expression of hypermethylated *BRCA1* in sporadic cancers is identical to the reduced expression in familial cancers where this gene is mutated[179,196,197]. It is important to note that *BRCA1* promoter hypermethylation is associated with a *BRCA1* tumour phenotype such as high grade, high mitotic count, little or no tubule formation, pushing margins, ER, PR and HER2 negativity [194,195]. *BRCA1* hypermethylation was also evident in the blood of some patients with *BRCA1* hypermethylated promoter in the tumour [194,195]. It is interesting to note that *BRCA1* has been shown to bind to the promoter region of DNMT1 and this binding is necessary for transcription of *DNMT1*. Loss of *BRCA1* expression as a result of promoter hypermethylation, leads to reduced *DNMT1* expression, which might be the cause of the global hypomethylation also evident in cancer cells [184].

Any gene can be silenced through hypermethylation, but silencing of certain genes can greatly affect carcinogenesis. These genes include tumour-suppresor genes such as *p16* and *cyclin dependent kinase inhibitor 2B (CDKN2B)* which regulate cell cycle[186,198,199], genes that suppress tumour invasion and metastasis, DNA repair genes such as *O6-methylguanine DNA methyltranferase (MGMT)*[179,200,201], hormone receptor genes, genes that inhibit angiogenesis, DNA damage response genes such as *BRCA1* [186,202,203] and mismatch repair genes such as MutL homolog 1 (*hMLH1)* and DNA mismatch repair protein (*hMSH2)*[186,204]. Aberrant methylation of the promoters of these genes leads to a loss of function, providing selective advantage to neoplastic cells [179]. Some tumours have the promoters of several of these genes methylated concurrently, producing what is known as the CpG island methylator phenotype (CIMP) which can be used as a marker for tumour progression [183].

The creators of PUBMETH, a database created in 2007, analysed the results of studies that investigated gene promoter hypermethylation and breast cancer. For each gene promoter, the number of references investigating the particular promoter and breast cancer, the number of samples analysed, and the average percentage hypermethylation are reported in the database. In the same way, I conducted a similar review

of the literature from 2007 up to May 2013 in order to identify any additional gene promoters, whose hypermethylation was associated with breast cancer. The search was conducted in PUBMED using the following search term: {"GENE PROMOTER HYPERMETHYLATION" OR "GENE PROMOTER HYPOMETHYLATION" AND "BREAST CANCER"}. All the genes identified from PUBMETH and from the literature review and the number of references in which each gene is investigated are listed in Appendix C. As shown in Appendix C, combining the data from PUBMETH with data from an updated literature review was crucial since a large number of genes has been investigated from 2007 onwards and thus were not included in the database. In the majority of these studies, hypermethylation of a gene promoter correlated with reduced expression of that gene.

One way in which methylation of promoter regions inhibits gene expression is that the usual transcription regulatory proteins cannot bind to the promoter sequence [205,206]. However, it has also been hypothesised that DNA methylation affects gene expression, by the recruitment of histone modifying proteins and a subsequent change in chromatin conformation. Generally euchromatin – the state that encourages gene transcription – is characterised by high levels of acetylation and trimethylated H3K4, H3K36 and H3K79 whereas heterochromatin – the state that is associated with gene repression – is characterised by low levels of acetylation and high levels of histone H3 lysine 9 (H3K9), histone H3 lysine 27 (H3K27) and histone H3 lysine 20 (H3K20) methylation [207]. One mechanism by which DNA methylation affects chromatin structure is that methyl binding proteins such as MeCP2 can recognize methylated regions on DNA and recruit histone deacetylases (HDACs) or histone methylases to the site[149,151,154,206,208–212]. Another possible mechanism is that HDACs interact with DNMTs directly and participate in gene silencing with or without accompanying DNA methylation[151,179,213–216]. Whichever the mechanism, the recruited HDACs remove acetyl groups from the ϵ-amino groups of lysine residues at the N-termini of histones and the recruited histone methylases, methylate histones at key amino acids. Histone deacetylation and methylation with the subsequent histone modification, cause chromatin to take up the heterochromatin conformation, thus obscuring access to transcription activating complexes, repressing transcriptional activity and silencing certain genes.

However, there is evidence that supports another alternative hypothesis, namely that it is chromatin structure and chemical modification of histones that defines DNA methylation[149,186,217]. Data has suggested that the pattern of methylation is dynamic, reversible, and dictated by chromatin structure. Firstly, activation of chromatin has been shown to bring about active replication-independent DNA methylation[178,216]. Secondly, inactivation of chromatin structure was shown to precede *de novo* methylation of the tumour suppressor gene *p16* in colorectal cancer cell lines[178,216]. It has been proposed that DNA methylation in somatic cells is maintained by a balance in methylation and demethylation activities, and this balance is determined by chromatin structure by recruiting DNMTs or gating accessibility to demethylases [178]. According to this hypothesis, regional hypermethylation results from local changes in

chromatin structure, which prevent access to demethylases [178] or chromatin repressive proteins, such as Polycomb Group (PcG) proteins, which can mediate methylation of histones at specific regions such as H3K27 and H3K9, thus marking target genes for DNA methylation and recruiting DNMTs and HDACs to the region[149,151,215,218–222]. At the same time, histone methyltransferases have been reported to direct DNA methylation to specific targets by recruiting DNMTs[207,223,224]. These views explain how regional hypermethylation can exist in the presence of global hypomethylation. On the other hand, methylation of H3K4 inhibits recruitment of DNMT3A by DNMT3L thus preventing *de novo* DNA methylation at specific sites[207,225].

Whichever the actual mechanism, epigenetic modifications have been shown to cause long-range gene silencing, with the aberrantly methylated regions being non-randomly distributed but concentrating in specific small genomic regions[111,226]. In addition to gene silencing, hypermethylation can render a region more unstable, thus predisposing to mutational events. Methylated cytosines are more prone to alterations by endogenous or exogenous factors that can lead to mutations, therefore, mutation rates at CpG sites have been estimated to be around 40 times higher than other transitional mutations [111,156]. 5-methyl Cytosine is more unstable than its unmethylated counterpart, since deamination of 5mC can lead to a G/T mismatch and to a C to T transition mutation [227,228]. In fact, one third of all known somatic and germ line mutations are C to T transitions at CpG nucleotides [111,229111]. As a result, in many human cancers sites of DNA methylation are "mutational hotspots." For example, 50% of inactivating point mutations in the coding region of the tumour suppressor *TP53* gene occur at methylated cytosines. Lastly, there is an association between areas of epigenetically and/or genetically disrupted genome and chromosomal deletions [179]. These deletions can cause loss of heterozygosity (LOH). For example, RASSF1A and hypermethylated in cancer-1 (HIC1) regions are characterized by frequent LOH in several tumours, and they are also found hypermethylated in many human cancers such as lung, prostate, colon, and breast[179,230–232].

As already alluded to above, methylation is a non-random process. Research seems to suggest that the genes that will be targeted for inactivation via hypermethylation, are encoded by intrinsic or extrinsic factors. It is not by chance that distinct tumour types show hypermethylation in several common genes but also in numerous different genes, displaying a unique profile of promoter methylation for each cancer type[149,153,233]. Also, some tumour types have more methylation of known CpG islands than other tumours – those of the GI tract being the most hypermethylated and ovarian cancers as well as sarcomas being the least methylated [155]. However, it is still not clear why some regions become hypermethylated while others are not. One explanation, as outlined above, is that chromatin structure and chromatin remodelling complexes, such as PcG proteins, direct methylation and demethylation activities to specific target regions. Secondly, it is possible that inactivation of a particular gene confers a growth advantage and is thus selected [207]. Thirdly it might be because, not all nucleotide sequences are equally prone to hypermethylation. Instead, some sequences are more prone to hypermethylation while others are

methylation resistant. Even CpG islands differ in their susceptibility to *de novo* methylation [149,234] therefore, some intrinsic factors direct methylation to specific target sequences. In fact studies suggest that there may be common sequence motifs in promoters that undergo CpG island hypermethylation [219,219,235]. It is also interesting to note, that most sequences prone to methylation are associated with CpG islands, whereas methylation resistant sequences are scattered along the genome in CpG island and non-CpG island DNA[149,236]. This explains why CpG islands in promoter regions are mostly hypermethylated in cancer, even when other CpG islands and the whole genome become unmethylated. Lastly, there is the possibility of the spreading of methylation from highly methylated sequences to their surroundings as suggested by Graff et al., 1997 [235,237]. Despite the many proposed hypotheses, our understanding of why only selected genes/regions become hypermethylated and the mechanisms by which they become so is limited.

1.3.1.3 Gene/Region Specific Hypomethylation

On the other hand, independently and less frequently than regional hypermethylation, CpG dinucleotides in other regions that are usually methylated may become unmethylated leading to poor repression of transcription, and increased expression of genes such as oncogenes[111,149,238149] or pro-metastatic genes [178]. Examples of pro-metastatic genes whose expression is regulated by DNA methylation in breast cancer are *MAGE*[149,239], *uPA*[178,240,241], *SINUCLEIN γ*[178,242,243], and *multidrug resistance 1 gene (MDR1*) [149,244].

In addition, regional hypomethylation can lead to loss of imprinting. A principal example is the maternally imprinted gene *Insulin like Growth Factor-2 (IGF-2)*. In maternal imprinting, the maternal copy of the gene is heavily methylated and thus not expressed, hence only the paternal copy is expressed. Consequently, imprinted genes are particularly vulnerable to environmental exposures since a) there is only one good copy expressed and mutations or expression changes on this copy could have detrimental effects on health and b) expression of both copies by loss of imprinting could lead to deleterious outcomes [245]. In the case of *IGF2*, loss of imprinting has been observed in many cancers including breast, colon, liver, and lung [207,246].

It would be expected that hypomethylation in the promoter region of a gene or in known enhancer sequences would affect gene expression more than hypomethylation within the gene. However, hypomethylation within gene regions might be coordinated with hypomethylation in upstream transcription regulatory sequences thus having an activating effect[166]. Lastly, hypomethylation of the normally hypermethylated allele of an imprinted gene can occur, and this has been shown to occur for *IGF2* in certain cancers[153,238,247,248]. However, the extent to which this might affect allele specific expression remains to be seen [166].

Recently, it was suggested that methylation changes in CpG island shores may also be involved in neoplasia[153,163,207,249,250] since hypermethylation in these sites can also lead to transcriptional repression.

These epigenetic changes not only lead to the development and/or progression of cancer, but they are also implicated in the type of breast cancer and in development of drug resistance. In animal models it has been shown that when undifferentiated breast cancer cells suffer prolonged exposure to oestrogen or oestrogen-like compounds, they are more likely to become neoplastic [127,151] or exhibit a cancer-like methylome [151,251]. This might be due to the hypomethylation and elevated expression of transcription factor genes involved in stem cell function in undifferentiated breast cells[151,252,253]. In addition, the loss of ER signalling in some cancers, and the associated increased aggressiveness of the tumour, have been linked to epigenetic silencing of ER-α target genes, via recruitment of PcG proteins and HDACs followed by DNA methylation in their promoter regions [151,254]. Lastly, it has been demonstrated that the development of anti-oestrogen resistance was associated with changes in methylation profiles of multiple genes [151,255], and resistance to tamoxifen was linked to epigenetic silencing of cyclin dependent kinase 10 (CDK10) [151,256].

Despite the wealth of knowledge gathered in the past few years regarding the epigenetic changes that occur in cancer cells, many questions concerning the origin and reason of these changes remain unanswered. However, several hypotheses regarding the cause of these epigenetic changes have been made. Firstly, mutations in genes coding epigenetic machinery can be implicated in the disturbed epigenome of cancer cells. To support this hypothesis, researchers have shown the presence of somatic mutations in *DNMT3a* in acute myeloid leukaemia [153,257]. Secondly, deregulation of *DNMT* expression – overexpression or downregulation – either due to changes in *DNMT* genes or due to miRNAs which can bind and deactivate the transcribed mRNA of DNMTs is possible [207,258]. Lastly, anomalous expression or activity of histone methyltransferases (HMTs) and histone demethylases (HDMs) due to chromosomal translocations, amplification, deletion, overexpression or silencing, has also been reported as a possible cause of the epigenetic changes associated with cancer [153]. Whichever the cause of the disturbed epigenome, one needs to always bear in mind the complex interactions between the epigenetic machinery as well as between the epigenetic changes. These interactions form what is best described as a complicated web rather than a clear cut cascade of events.

Epigenetic changes and especially DNA methylation have been the focus of much attention not only because it can help researchers understand more about the processes involved in carcinogenesis but also because it constitutes an attractive candidate as a biomarker for early detection, therapy monitoring, assessment of prognosis, prediction of therapy response and even treatment[259].

1.3.2 Epigenetics as Biomarker for Early Detection/Diagnosis

Mammography screening has been used for the early detection and diagnosis of breast cancer but its sensitivity and specificity are not perfect. The sensitivity varies due to variations in practitioners' experience and skill and it is reduced in younger women, women of Asian descent, women on HRT and those with denser breasts [260–264]. Also, specificity is not perfect as false positive results leading to unnecessary biopsies and anxiety are common [264–266]. Hypermethylation of promoter regions can be a useful tool in improving the detection of all types of breast cancer and separating invasive malignant tumours from benign breast disease. Increased sensitivity and specificity would reduce false negative rates, helping to catch more cancers at an early phase, but also false positive rates, helping to minimize the number of unnecessary biopsies and reduce anxiety.

DNA methylation and more specifically DNA sites that are frequently hypermethylated in cancer but are always unmethylated in normal cells have been proposed as a marker for early detection of breast cancer. Methylation patterns either at the promoters of genes or at various methylation variable regions and CpG dinucleotides along genes have been used for this purpose (reviewed in [264]).

For example, Jeronimo et al. (2008) [267] evidenced that a panel of four genes (*CCND2, RASSF1A, APC, and HIN1*) from tissue DNA was able to distinguish between invasive carcinomas, fibroadenomas, and normal tissue. However, in general, the studies investigating promoter hypermethylation and its ability to distinguish between benign breast disease (BBD) and malignant tumours have yielded mixed results. Dulaimi et al. (2004)[268] have evidenced that promoter hypermethylation of three genes (*APC, RASSF1A, DAPK*) was present in DNA from in situ and invasive carcinomas but absent from BBD patients and controls. On the other hand, Parella et al.(2004)[269] investigated several genes including *BRCA1, p16INK4A, ESR-1, GSTP1, TRβ1, HIC1, APC, CCND2 and CDH1*, and found that promoter hypomethylation in fibroadenomas was similar to that in invasive disease, with the exception of *CDH1* which was more specific. Another panel of 4 genes, namely *RASSF1A, RARβ2, APC, and CCND2* , was used to evidence that breast tissue methylation levels from patients with BBD were in between methylation levels in breast tumour tissue and methylation levels in healthy breast tissue from controls [270]. Lastly, when investigating three genes (*RASSF1A, RARβ2 and CCND2*), Pu et al. (2003) found that the frequency of promoter methylation increased from benign, to in situ carcinoma, and invasive carcinoma [271]. In addition, studies developed and tested panels of 3,4, and 9 genes with improved sensitivity to cytological analysis [151,272–274]. These conflicting results indicate the need for further research.

The need for more research is further enhanced by the promise held by initial results. For example, promoter hypermethylation can also help improve the detection of Invasive Lobular Carcinoma (ILC) which is not easily detected by mammography, and also distinguish ILC from Invasive Ductal Carcinoma (IDC)

[264,271,273,275]. The great promise held by promoter hypermethylation as a marker of carcinogenesis is reviewed in Kelly et al. (2010) [276] and Brooks et al. (2009) [264]. One striking example is *BRCA1* promoter hypermethylation [170,195,277]. Results of these studies suggest that *BRCA1* and other genes' promoter hypermethylation in the tumour is usually also evident in peripheral blood. Despite the fact that not all studies demonstrated that promoter hypermethylation of certain genes can distinguish blood from patients and controls [170,195,277], other studies have identified distinct methylation patterns in the blood of cancer patients compared to that of controls [278–280].

Of great importance are the findings regarding methylation of the ATM gene. Brennan et al. (2012) [281] used prediagnostic blood samples from three studies to analyse genomic DNA methylation of two ATM intragenic loci. The results indicated that highest methylation at an intragenic locus on the *ATM* gene was associated with an increased risk of breast cancer in women, and the blood of cases was collected on average 18-65 months prior to diagnosis. Thus gene body hypermethylation[282] as well as global hypomethylation [170] in peripheral blood are promising and least invasive potential markers for early diagnosis.

Circulating tumour DNA methylation holds also promise as a biomarker for early detection since the blood of breast cancer patients contains detectable amounts of methylated DNA [151,199,283,284]. Interestingly, 93% of the circulating DNA is derived from the tumour [264,285] and there is good concordance between tumour and blood samples from the same patients [264]. Methylated DNA is released early in the disease process, even before the cancer becomes invasive [154,264], and methylation accumulates as the disease progresses [264,286]. However, the sensitivity of gene panels obtained from circulating DNA is slightly lower than that of DNA from tumours [151].

1.3.3 Epigenetics as Biomarker for Environmentally Induced Risk

Methylation in peripheral blood DNA can be a useful biomarker not only of the presence of disease but also a marker of environmentally induced cancer risk. This is because blood is affected not only by internal exposures but also by external exposures such as lifestyle choices like smoking, diet etc. In addition, given the role of blood tissue in immune responses, investigation of blood DNA methylation might reflect immune responses to external and internal stimuli [169]. In fact, several environmental exposures have been shown to impact on DNA methylation [287] which can act as a reflection of certain exposures. For example, folate and other one-carbon metabolism nutrient intake has been shown to impact on DNA methylation [288–290], and smoking habits were recently evidenced to leave their mark on the methylation pattern of a cell [291]. Alcohol drinking was also shown to be associated with peripheral blood methylation [292]. However, several caveats need to be considered when investigating DNA methylation in peripheral blood. Firstly, blood is a very heterogeneous tissue and DNA methylation is very tissue specific [146,169,293]. Differential methylation in the different blood cell types and the inevitable "averaging" when blood methylation is investigated may be masking the epigenetic effects of cancer or environmental exposures [169,292]. In addition, presence of undetected or asymptomatic disease may be impacting on blood methylation. For example, disease may be causing increased leukocyte proliferation which might lead to hypomethylation if there are reduced substrate, energy, and machinery resources [169]. This issue is very difficult to resolve given the unknown latency periods of many cancers.

1.3.4 Epigenetics as Prognostic Tool for Response to Treatment

Epigenetics can be used not only as a marker for early detection but also as a prognostic tool for response to treatment. This is because the drugs used for chemotherapy could target genes that are epigenetically silenced in tumour cells, conferring a survival advantage to the tumour cells [151]. For example, tumours that lack ER-α expression have been shown to be insensitive to endocrine therapy [151,294,295] and *CDK10* methylation can serve as a novel marker for determining a patient's response to therapy as low levels of *CDK10* expression are associated with early relapse under tamoxifen treatment [153,256,296]. In addition, *paired like homeodomein transcription factor 2 (PITX2)* DNA methylation was validated for clinically relevant outcome prediction in early breast cancer patients on tamoxifen [259]. Given that methylation patterns are reversible, demethylating agents can be potentially used to desensitise resistant cancer cells to chemotherapy agents [297] but this requires further research and exploration (reviewed in[153]).

On the other hand, DNA repair genes when active have the potential of repairing DNA damage induced by many chemotherapy agents, generating chemoresistance. However, if hypermethylated and silenced then chemotherapy is likely to be more effective [298,299]. DNA repair genes that have the potential to be used as markers to the response to treatment include *MLH1, BRCA1, GSTP1,* and *MGMT* in the treatment of ovarian cancer [299,300], and *BRCA1* for PARP inhibitors treatment in breast cancer[301].

Understanding the available knowledge pool and the mechanisms involved in therapy response has led to the identification of DNA methylation patterns associated with progression free survival in ovarian cancer [302] and the principle can be extended to other tumours, as well.

However, key in these investigations remains the need to determine those DNA methylation changes that drive resistance in contrast to those changes that occur either as a result of chance or as a consequence of the causative DNA methylation changes [303]. Only such discrimination will enable successful and efficient use of epigenetics as biomarkers for response to treatment.

1.3.5 Epigenetics as Potential Cancer Treatment

Due to the fact that in contrast to genetic alterations, epigenetic alterations are mostly reversible, they have the potential to be manipulated as part of cancer treatment. In fact, two DNMT inhibitor drugs and two Histone DeAcetylase Inhibitor Drugs have been approved for use by the Food and Drug Administration in the US [153].

DNMTs inhibitors work by incorporating themselves into DNA at the place of cytosine residues during replication from where they can covalently bind to and trap DNMTs [151,304,305]. However, due to their lack of specificity they have toxic side effects. On the other hand, non-nucleotide DNMT inhibitors have the ability to bind to the active sites of DNMTs and prevent successful interactions between the enzymes and their target sites [151,306–309].

1.4 DNA Methylation and Features Undergoing Secular Changes

It is established that much of the variation in epigenetic patterns is heritable and can be explained by common genetic variation. This genetically determined epigenetic variation is termed obligatory epigenomic variation[310]. Several studies have provided evidence for obligatory epigenomic variation and are reviewed by French et al [310]. However, there is also ample evidence that the heritability of DNA methylation is not 100%, suggesting non-genetic influences[311]. In fact, numerous environmental exposures have been implicated in contributing to epigenome variation [183]. Existing literature, although not always consistent, demonstrates that, in addition to other environmental exposures, the factors exhibiting secular trends may also leave their mark on DNA methylation.

1.4.1 Socioeconomic Status and DNA Methylation

Several studies have investigated the impact of social and economic conditions on the epigenome (Table 1). These studies were identified as the most relevant from a systematic literature review up to May 1st, 2013, using the following two search terms: 1. ("social class"[MeSH Terms] OR ("social"[All Fields] AND "class"[All Fields]) OR "social class"[All Fields] OR ("socioeconomic"[All Fields] AND "status"[All Fields]) OR "socioeconomic status"[All Fields]) AND ("dna"[MeSH Terms] OR "dna"[All Fields]) AND ("methylation"[MeSH Terms] OR "methylation"[All Fields]) and 2. ("education"[Subheading] OR "education"[All Fields] OR "educational status"[MeSH Terms] OR ("educational"[All Fields] AND "status"[All Fields]) OR "educational status"[All Fields] OR "education"[All Fields] OR "education"[MeSH Terms]) AND ("dna"[MeSH Terms] OR "dna"[All Fields]) AND ("methylation"[MeSH Terms] OR "methylation"[All Fields]) in PUBMED. The cited references of the identified studies were also searched for additional relevant publications. Two hundred and sixty nine studies were retrieved, only 8 of which were relevant: investigating measures of SES and DNA methylation in human subjects. These relevant studies are presented in Table 1.

Most studies show that early life social conditions leave a mark on the epigenome. Early life socioeconomic position (SEP), a variable reflecting a combination of father's occupation and access to household amenities, was shown to cause differential methylation of 1252 promoters [312]. Similarly, adult global hypomethylation was evidenced in the most socioeconomically deprived at birth subjects of the pSoBid cohort [313]. As already discussed, epigenetic responses to early life conditions are time and stage-specific. Heijmans et al. (2008) demonstrated that periconceptional – but not late gestational – exposure to famine was associated with persistent hypomethylation of CpG loci in the IGF2 differentially methylated region [142]. In the same study, late gestational – but not periconceptional – exposure to famine was associated with lower birth weight. To further support the time-specificity of adult DNA methylation changes in response to

early life conditions, the investigation of 15 candidate genes involved in metabolism, CVD, and growth in subjects exposed and not exposed to famine, demonstrated that different genes were differentially methylated depending on the timing of exposure (periconceptional vs. late gestational exposure) [147]. Lastly, Lam et al. (2012) [314], demonstrated that the pattern of methylation in low early life SES individuals differed from that expected by chance. However, only 3 CpGs had significantly different methylation between different early life SES groups, and the difference between the groups was <5% thus with questionable biological significance. Using expression data to support the influence of early life socio-economic conditions on adult biology, Miller et al. (2009) [315], evidenced up-regulation of genes involved in inflammatory responses in adulthood, in subjects spending their early life in unfavourable socioeconomic conditions, and this association was independent of adult SES, lifestyle practices, and perceived stress [315].

On the other hand, investigation of global DNA methylation in response to pre-natal nutrition, using surrogate measures of global methylation such as LINE1 and Sat2 methylation, did not yield any significant results [316]. Similarly, family socioeconomic status was not significantly associated with adult global DNA methylation in a multi-ethnic birth cohort in New York [317].

Adult SES was also shown to be associated with DNA methylation. Borghol et al. (2012) [312] demonstrated that mid-adulthood SEP was associated with aberrant methylation of 545 promoters investigated. In another prospective investigation, household's highest occupational position significantly influenced methylation at probes associated with pro-inflammatory genes, with low SEP predicting lower methylation [131]. In a New-York based cohort, while income, education, and childhood SES did not predict LINE-1 methylation, a one standard deviation higher wealth was associated with 0.09% higher Alu and 0.15% lower LINE-1 methylation, and the associations remained statistically significant after adjustments for age, gender, and race/ethnicity [318]. However, these effect estimates are not likely to denote changes with biological significance.

On the contrary, adult SES was not associated with methylation differences in a community based cohort, however, adult total cortisol levels, a marker of adult stress, were associated with small methylation differences in 5 CpGs [314].Similarly, education was not associated with global DNA methylation in peripheral blood in another investigation [171].

Table 1 – Studies investigating the association between Socioeconomic Status and DNA methylation

First author, year	Subjects, N	Sample type	Exposure	Outcome	Analysis and Effect Estimate	P-value	Confounders
Heijmans, 2008[142]	Early life famine exposed adults and same-sex siblings, 244	Peripheral Blood DNA	Periconceptional exposure to famine, n=60 Late gestational exposure, n=62 Unexposed same sex siblings, n=122	*IGFR2* DMR∝ methylation using a mass spectrometry–based method	Linear mixed model comparing periconceptional exposure to no exposure. Effect estimates are relative change in the exposed: Average *IGFR2* DMR: -5.2% Linear mixed model comparing late gestational exposure to no exposure. Effect estimates are relative change in the exposed: Average *IGFR2* DMR: -0.9% Interaction between timing and exposure	$5.9x10^{-5}$ 0.64 $4.7x10^{-3}$	Age Age
Terry, 2008[317]	Adult women, 92	Peripheral Blood DNA	Family SES∝ (continuous)	Genomic DNA methylation using the [H^3]-methyl acceptance assay	Linear regression – outcome: methylation, exposure: age at menarche (continuous) β=-0.01 (-0.01-0.002)	NA∝	Prenatal smoke exposure, race, family SES, birth weight, birth length, age at menarche, childhood passive smoke exposure, smoking status, parity, age at first birth
Tobi, 2009[147]	Early life famine exposed adults and same-sex siblings, 244	Peripheral Blood DNA	Periconceptional exposure to famine, n=60 Late pregnancy exposure, n=62 Unexposed same sex siblings, n=122	Methylation of 15 loci involved in metabolic and cardiovascular disease using a mass spectrometry–based method: *GNASAS, GNAS A/B, MEG3, KCNQ1OT1, INSIGF, GRB10, IGFR2, IL10, TNF, ABCA1, APOC1, FTO, LEP,* *NR3C1* and *CRH*	Linear mixed model comparing periconceptional exposure to no exposure. Effect estimates are within-pair differences divided by the standard deviation in the sibling controls. *INSIGF*: -0.61 *IL10:* 0.37 *GNASAS*: 0.24 *LEP:* 0.24 *MEG3*: 0.21 *ABCA1* meth: 0.17 Linear mixed model comparing late pregnancy exposure to no exposure. Effect estimates are within-pair differences divided by the standard	$2.3x10^{-5}$ $1.8x10^{-6}$ $3.1x10^{-6}$ $2.9x10^{-3}$ $8.0x10^{-3}$ $8.2x10^{-4}$	Family relations, bisulphite batch and age at blood draw

					deviation in the sibling controls. *GNASAS*: -0.26	$1.1x10^{-7}$	
Borghol, 2011[312]	Adults, 40	Peripheral Blood DNA	Childhood SEP$^{\propto}$ and Adulthood SEP	Methylation of 223359 probes (mapping to 20000 genes and 400microRNAs) using Methylated DNA Immunoprecipitation	For childhood SEP: association with 1252 gene promoters enriched in the following functional categories: Extra-cellular signalling, Intra-cellular signalling, DNA signalling, and Metabolism For adulthood SEP: association with 545 gene promoters	NA	NA
Lumey, 2011[316]	Adults, 947	Peripheral Blood DNA	Prenatal exposure to famine, n=350 Unexposed time controls, n=290 Unexposed same-sex siblings, n=307	Genomic DNA methylation – LINE 1 and Sat2 using pyrosequencing Global DNA methylation using LUMA	Linear model comparing exposed to all non-exposed. Effect estimates are relative change in the exposed: **LUMA:** 0.16 (-0.49 – 0.81) **LINE-1:** -0.05 (-0.33 – 0.22) **Sat2:** -0.51 (-7.38 – 6.36)	0.63 0.70 0.88	Age at examination and within family clustering
McGuiness, 2012[313]	Subjects of the pSoBid cohort, 239	Peripheral Blood DNA	Deprivation group: affluent vs. deprived Social class: manual vs. non manual Years of Education	Global DNA methylation using Methylamp kit	% difference in methylation predicted from linear regression models: **Deprived vs. affluent:** -17.3 **Manual vs. non-manual:** -27.4 **Years of education:** 2.5	0.015 0.003 0.052	Income, diet, smoking Age, gender, deprivation group, income, diet, smoking Age, gender, deprivation group, income, diet, smoking
Lam, 2012[314]	Adults, 92	Peripheral Blood DNA	Early-life SES: Low SES (n=46) vs. high SES (n=46) Current SES: Low SES (n=46) vs. high SES (n=46)	Genome-wide methylation at ~27k CpG sites using the Illumina HumanMethylation 27 array platform	Early life SES: 3 differentially methylated CpGs & a non-random P-value distribution Current SES: 0 differentially methylated CpGs & a P-value distribution as expected by chance	NA	NA
Subramanyam, 2013[318]	Adults, 987	Peripheral Blood DNA	Income (continuous) Wealth (continuous) Education (continuous)	Genomic DNA methylation – LINE 1 and Alu using	Mean differences in DNA methylation per 1 SD change in the exposure variable:		

			Childhood SES (continuous)	pyrosequencing	**LINE-1:** Income: 0.02 Wealth: -0.12 Education: 0.04 Childhood SES: 0.04 **Alu:** Income: -0.037 Wealth: 0.08 Education: 0.01 Childhood SES: 0.02	 0.06 0.05 0.06 0.06 0.035 0.03 0.04 0.04	Age, gender, and race

∝ SES, socioeconomic status; SEP, socioeconomic position; DMR, differentially methylated region; NA, not available

1.4.2 Nutrition and DNA methylation

The evidence regarding the impact of nutrition on DNA methylation is controversial depending on the aspect of diet examined. Several studies, especially on folate, alcohol and other one carbon metabolism involved nutrients, suggest that some dietary exposures can impact on DNA methylation. One-carbon metabolism is the pathway via which methyl groups are generated and added to DNA/RNA to confer methylation or are used for the synthesis of purines. Changes or inadequacies in this pathway have been linked to carcinogenesis [319,320].

1.4.2.1 Folate, alcohol, one-carbon metabolism and DNA methylation

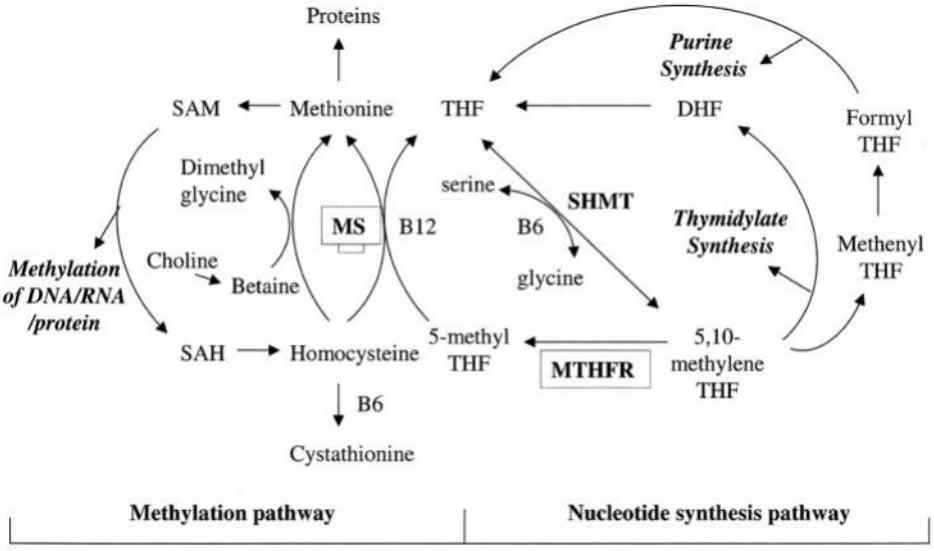


Figure 11 – Folate in one-carbon metabolism.

MTHFR sits as a fulcrum between DNA methylation and DNA synthesis, balancing the two pathways to maintain normal homeostasis. DHF: dihydrofolate; THF: tetrahydrofolate; MS: methionine synthase; MTHFR: Methylene tetrahydrofolate reductase; SAH: S-adenosylhomocysteine; SAML S-adenosylmethionine; SHMT: serine hydroxymethyl transferase. Reproduced with permission from [290].

Folate in Nucleic acid Metabolism: Folate exists in many co-enzymatic forms in mammalian cells, but the role of all of these forms is the transfer of one-carbon units. Figure 11 describes the pathway for this transfer and shows how the various forms of folate are implicated. The key step in the transfer of one-carbon units is the methylation of homocysteine to methionine. This transfer is catalysed by 5-methyltetrahydrofolate:homocysteine methyltransferase, and Methyl Tetrahydrofolate (methylTHF) serves as the substrate, donating its methyl group to homocysteine. Vitamin B12 is a necessary co-enzyme for the reaction. Methionine is then converted to S-adenosylmethionine (SAdoMet) by methionine adenosyl transferase, and SAdoMet can donate its methyl group which it derived from 5 methylTHF for biological methylation reactions and *de novo* deoxynucleotide triphosphate synthesis. The transfer of the methyl group from SAdoMet to the 5 carbon of the cytosine ring is catalysed by DNMT. In liver and kidney cells, methionine can also be produced in an alternative pathway that is not dependent

on folate, the methylation of homocysteine by betaine. Despite the existence of this alternative pathway, folate depletion can sufficiently diminish SAdoMet pools in the cell [109,228,290] and result in diminished levels of methylation [321].

Depletion of folate not only affects methylation of biological molecules such as DNA and RNA, but it also adversely affects DNA synthesis and repair. As stated, folate derived one carbon groups are essential in the *de novo* synthesis of the pyrimidine, thymidylate, as well as the purines. The synthesis of thymidylate is a rate limiting step for DNA synthesis and requires 5,10-methyleneTHF as a coenzyme. When folate levels are adequate 5-methylTHF is used as a substrate for the production of SAdoMet and 5,10-methyleneTHF is used for dNTP synthesis. However, in folate depletion the mechanisms compete. If SAdoMet concentration in the cell decreases, the cell increases the conversion of 5,10-methyleneTHF to 5-methylTHF by the enzyme Methylenetetrahydrofolate reductase to compensate for the decrease, but this compromises the availability of 5,10-methyleneTHF for *de novo* dNTP synthesis. Since dNTPs are the immediate substrates for enzymes involved in DNA replication and repair, the reliability of these processes depends on the correct balance and availability of deoxynucleotides. Inadequate folate and an upset balance or unavailability of dNTPs will compromise these processes [228,290].

One proposed mechanism for folate associated carcinogenesis is altered DNA methylation via disturbances of the one-carbon metabolism. DNA methylation is not aberrantly altered in healthy individuals with adequate folate [322–324]. However, folate depletion can cause genome-wide hypomethylation and it can cause both hyper and hypomethylation of specific gene regions thus affecting carcinogenesis. Firstly, folate depletion and SAdoMet decreased concentrations can lead to decreased methylation in the whole genome, and this is nearly always observed in carcinogenesis [228,325]. Such a decrease in genomic methylation appears early in carcinogenesis and seems to precede mutation and deletion events[326]. Different studies have reported diminished genomic DNA methylation in individuals with folate depletion [288–290] but most published evidence concerns colorectal cancer. In addition it was shown that folate therapy restores DNA methylation in individuals whose cells are hypomethylated [288–290,323] but daily folic acid supplementation for 12 weeks did not impact on methylation in other studies on healthy individuals with normal folate status [322,327]. Secondly, folate depletion has been shown to induce hypomethylation of the coding regions of some genes, including *TP53* [328,329]. Cell culture studies show that aberrant methylation within the coding regions of genes can initiate mutation [330] and may induce breakage of the DNA backbone [331]. These breaks are due to increased uracil insertion, or deamination of unmethylated cytosines which is most common when SAdoMet concentrations are low [290,332,333]. Hypomethylation of promoter regions can also lead to increased proto-oncogene expression[324]. On the other hand, it has been shown in animals that folate depletion associated hypomethylation can be followed by a rebound hypermethylation at a later time [330,331,334]. It is hypothesised that this is due to the damaging effects of folate deficiency which moves DNMT, also a repair protein, to regions of folate deficiency induced damage and away from the DNA replication fork. This would result to regional hypermethylation, despite the global hypomethylation [335]. Despite these results in folate

depleted individuals, the results in folate deficient individuals have been varying according to treatment regime, tissue, and genes examined (reviewed by Kim Y-I [324]).

Alcohol in Nucleic acid Metabolism and Effects on Carcinogenesis: As already discussed, alcohol can also adversely affect one-carbon metabolism and its effects could be exerted through the following pathways (reviewed in [336] and demonstrated in Figure 12):

Decreased folate bioavailability

Heavy alcohol drinkers have been shown to have inadequate folate in their diet [337]. In addition, alcohol causes malabsorption of folate due to loss of deconjugation – necessary activity at the brush border – and partial loss of both folate binding proteins [338,339]. Also, alcohol causes increased renal excretion of folate due to a reduction in tubular reabsorption [340]. Lastly, superoxide radicals produced during the oxidation of acetaldehyde (the major metabolite of alcohol) can chemically destroy folate molecules in vitro [341].

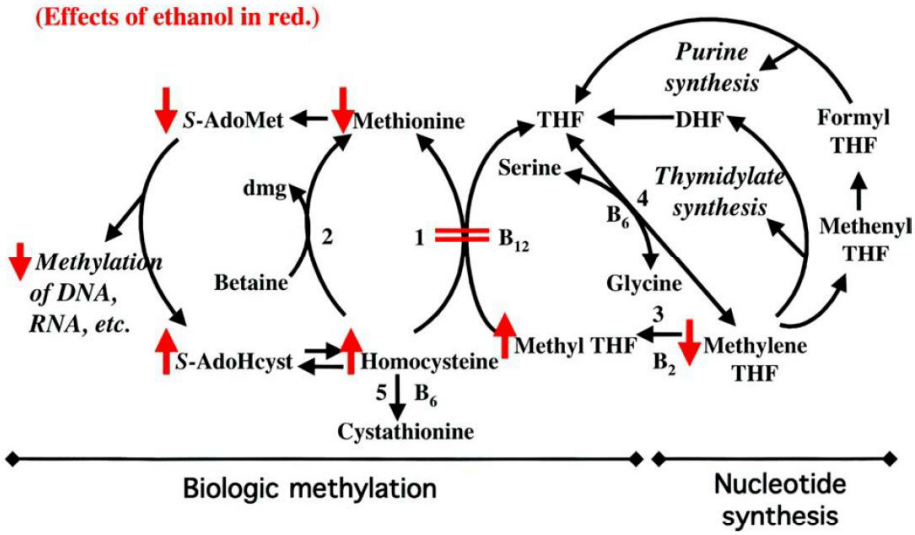


Figure 12 – The predominant effect of ethanol on one-carbon metabolism appears to be the inhibition of methionine synthase.

DHF: Dihydrofolate; dmg: dimethylglycine; S-AdoHcyst: S-adenosylhomocysteine; S-AdoMet: S-adenosylmethionine; THF: tetrahydrofolate; 1: methionine synthase; 2: betaine:homocysteine methyltransferase; 3: methylenetetrahydrofolate reductase; 4: serine hydroxymethyltransferase; 5: cystathionine b-synthase. Reproduced with permission from [336].

Adverse effects on folate dependent intermediary metabolism

One of the crucial effects of alcohol on the one-carbon metabolism is the inhibition of methionine synthase (MS) [342-343] and thus obstruction of methylation of homocysteine to give methionine. One hypothesis for MS inactivation is that acetaldehyde produces adducts with the MS protein. Also, alcohol can diminish serum Vitamin B12 in postmenopausal women [344] and B12 is an important co-factor for MS. Whichever the mechanism of MS inactivation, there is reduced methionine and SAdoMet in the cell and increased homocysteine and S-adenosylhomocysteine (homocysteine levels can rise only by a small percentage as some homocysteine can be

converted to S-AdoHcyst directly in a reversible reaction) [342]. Due to the decreased availability of SAdoMet, there is reduced capacity for methylation of DNA, since it is SAdoMet that donates its labile group for methylation. Also S-AdoHcyst is an inhibitor of all methyltransferases further reducing methylation capacity [345].

This may lead to global hypomethylation or altered methylation in specific genes or loci [179]. Even though it seems paradoxical, alcohol and folate deficiency may cause hypermethylation as well [346], since there seems to be a compensatory up-regulation of DNA methyltransferase activity [331] following extended periods of folate deficiency. In fact, alcohol has been shown to decrease expression of E-Cadherin [347], and this effect could be through hypermethylating its promoter since the E-Cadherin gene is frequently found to be hypermethylated in breast cancer [348]. A case-control study has shown that higher alcohol intake is associated with hypermethylation of the promoter of the E-cadherin gene in breast tumours and the effects are more pronounced for ER- tumours [346]. Lastly, it was shown that DNA methylation of E-Cadherin and p16 is associated with drinking intensity, independently of total consumption [346]. In a different case control study on African American women, it was shown that alcohol intake is associated with increased promoter methylation of the *ERα* gene in breast tumours [349]. These results demonstrate that the timing and pattern of exposure are important in determining alcohol intake's effect on the epigenome, and alcohol is associated with both hypo- and hyper-methylation events. Despite these promising finding, these case-control studies are lacking because one cannot discount any influence of disease.

Another effect of alcohol induced MS inactivation is an accumulation of 5 methyltetrahydrofolate which cannot be used for any other purpose, as opposed to other forms and this is termed as the *methylfolate trap*. Besides, SAdoMet is an inhibitor of MTHFR and in its absence more 5,10-methyleneTHF is converted to 5methylTHF. When a *methylfolate trap* develops there is loss of methylene- methenyl- and formyltetrahydrofolates leading to a decrease in purine and thymidine synthesis. As a result, DNA synthesis is defective, there is mis-incorporation of uridine in the place of thymidine, and attempts to correct the error leads to strand breaks[350].

Alcohol can also decrease the transcription of several other enzymes in one-carbon metabolism like MTHFR, methionine adenosyltransferase 1A, glycine N-methyltransferase, S-AdoHcyst hydrolase[351], further affecting the maintenance of a healthy methylation state. However, these effects have only been evidenced in the liver[351].

Alcohol might also increase serine synthesis since its inhibition of MS leads to an increase in homocysteine levels and a decrease in methionine availability. Serine synthesis may be increased by up-regulation of the expression of serine hydromethyltransferase enzyme. However, this would further deplete 5,10-methyleneTHF and would further reduce thymidine synthesis. This has not been evidenced directly but supporting evidence exists [352].

However, there are some agents that might attenuate the adverse effects of ethanol on folate metabolism. An alternative pathway exists for the remethylation of homocysteine that does not depend on folate or B12 and is therefore not susceptible to alcohol inhibition. In this pathway, a methyl group is transferred from betaine to homocysteine by betaine-homocysteine methyltransferases. This forms methionine and dimethylglycine. Betaine

supplementation in animals has helped attenuate the adverse effects of alcohol [353–355] while choline has not been as effective [356], but again all these effects are evident in liver cells.

Given the dependency of the one-carbon metabolism on an adequate supply of folate and the aforementioned effects of alcohol on the one-carbon metabolism, and given the crucial role of the one-carbon metabolism in DNA methylation, it is not surprising that altered levels of involved nutrients due to the nutrition transition can result in aberrant DNA methylation.

On the other hand, the impact of other foods/nutrients, shifting during the nutrition transition, such as fats and sugars, on DNA methylation is not expected to be great since they are not involved in mechanisms influencing DNA methylation.

An extensive literature review on the impact of dietary and alcohol intakes on DNA methylation in humans was conducted in PUBMED to identify relevant publications up to May 1st, 2013. The following search term was used: {DIET[Mesh] OR NUTRIENTS[Mesh] OR FOODS[Mesh] OR FOLATE[Mesh] AND METHYLATION[Mesh]} and the cited references in identified publications were further searched for relevant studies. The publications investigating methylation in human tissues were grouped according to outcome – GENE-SPECIFIC METHYLATION and GLOBAL METHYLATION. The results of the literature search are presented in Figure 13.

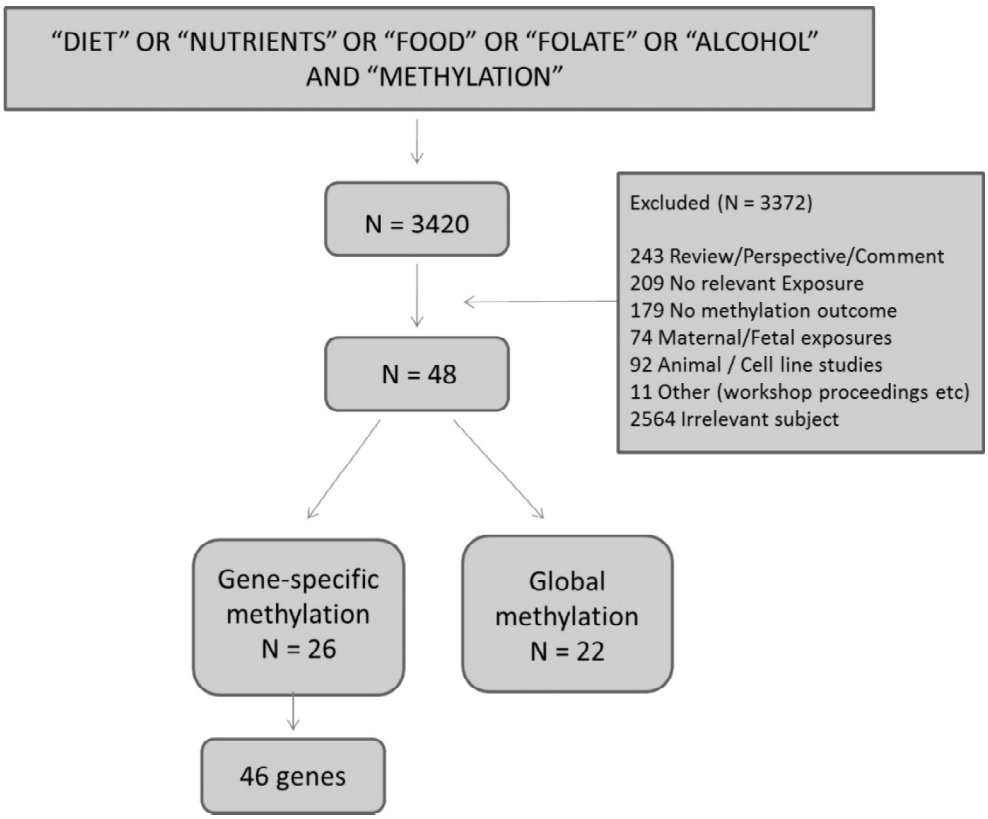


Figure 13 – Literature review on dietary factors and DNA methylation: Flow Chart

The identified relevant papers for each outcome are presented in APPENDIX D:Tables 1-2, sorted by exposure of interest and year of publication.

The heterogeneity in study designs, exposures, platforms/techniques, and outcomes did not permit a formal meta-analysis to be carried out. However, as expected, several, but not all, studies show that folate and alcohol intakes impact on the epigenome – both on global and gene specific methylation. The studies investigating measures of folate exposure – dietary intake or folate levels in biological samples, are diagrammatically presented in Figure 14. Each box represents an association investigated in the study with the respective first author and year of publication. The majority of investigations (n=15) show a positive association between measures of folate exposure and global DNA methylation (indicated by the green colour). Still however, 13 investigations evidenced non-significant associations (yellow colour) and three studies evidenced significant inverse associations between folate exposure and global DNA methylation (Figure 14). The studies investigating B-vitamin and alcohol intake did not evidence a significant effect on global DNA methylation but a few of the studies demonstrated associations with gene specific methylation. Lastly, there is very little to no evidence on other nutritional exposures such as caloric intake, micronutrients, fats and sugars (APPENDIX D).

From a careful examination, it becomes evident that several of the genes found to be aberrantly methylated in response to dietary exposures (Table 1, Appendix D), were also found to be aberrantly methylated in breast cancer (APPENDIX C). These genes include *p16(INK4)*, *RASSF1A*, *hMLH1*, *ESR1*, *RARβ2*, *O6-MGMT*, *ERα, APC, TP53*, *GSTP1*, *Socs1*, *SFRP1*, *Rassf1a, CCND2, IGF2, RUNX3, EGFR, TIMP3* and *ARF*. This commonality lends support to the hypothesis that nutritional changes can impact on breast cancer risk, through DNA methylation and suggests that methylation might be a valid biomarker for the investigation of the relationship between diet and breast cancer.

Figure 14 – Investigations in studies of the association between measures of folate exposure and global DNA methylation.

Each box represents an association. Some studies investigated more than one associations (different measures of exposure). Boxes in green indicate significant positive associations. Boxes in red indicate significant inverse associations and yellow boxes signify non-significant associations.

Folate measurements from biological samples	Folate intake
Perng (2012)	Ono (2013)
Wang (2012)	Crider (2012)
Figueiredo (2009)	Crider (2012)
Figueiredo (2009)	Crider (2012)
Pilsner (2007)	Crider (2012)
Pufulete (2005)	Crider (2012)
Friso (2002)	Jung (2011)
Piyathilake (2000)	Elworthy (2011)
Piyathilake (2000)	Christensen (2010)
Fowler (1998)	Schernhammer (2010)
Fenech (1998)	Figueiredo (2009)
Cravo (1994)	Piyathilake (2006)
	Basten (2006)
	Pufulete (2005)
	Rampersaud (2000)
	Rampersaud (2000)
	Fenech (1998)
	Jacob (1998)
	Cravo (1994)

1.4.3 Overweight/Obesity and DNA Methylation

Several studies have investigated the association between overweight/obesity and DNA methylation, using various study designs and different variations of the exposure and outcome. To identify these studies, an extensive literature review in PUBMED was conducted up to May 1st, 2013 using the search term: ((("overweight"[MeSH Terms] OR "overweight"[All Fields]) OR ("obesity"[MeSH Terms] OR "obesity"[All Fields])) OR "bmi"[All Fields])) AND ("dna"[MeSH Terms] OR "dna"[All Fields]) AND ("methylation"[MeSH Terms] OR "methylation"[All Fields]). Four hundred and sixty studies were retrieved, only some of which were considered relevant: investigating the association between overweight/obesity and DNA methylation in human tissues. The references of the relevant studies were searched to identify any additional publications on the subject matter. In total, 19 relevant studies were identified and presented in Table 2.

Associations between overweight/obesity and DNA methylation have been demonstrated in infants, children, adolescents, and adults. Newborns with low or high birth weight were shown to have significantly but modestly lower cord-blood LINE-1 methylation compared to normal-weight infants, and this association remained even after adjustment for relevant confounders such as gestational age, sex, maternal age at delivery and maternal smoking during pregnancy [357]. Methylation status of two CpG islands on the *leptin-responsive gene preproopiomelanocortin* (*POMC*), which was previously shown to be associated with obesity, was investigated in children's peripheral blood DNA [358]. One of the two CpG islands investigated was found to be hypermethylated in obese compared to normal weight children[358]. In adolescents, genome wide methylation analysis in fasting blood samples revealed two CpG loci differentially methylated between obese and lean individuals. One CpG site on the *UBASH3A* gene was hypermethylated in obese compared to lean subjects, whereas the second CpG site on the *TRIM3* gene was hypomethylated in obese adolescents[359]. Also in 17-year-olds, methylation at the *IGF2/H19* imprinted region was shown to be significantly and positively correlated with skin fold thickness and subcutaneous adiposity, but not with weight, height, BMI, waist circumference or visceral adiposity[360]. In adults, methylation analysis in white blood cells (WBC) DNA revealed significant hypermethylation in CpG sites of the *CLOCK* and *BMAL1* genes between normal-weight and overweight/obese individuals. CpG sites in *PER2*, a third gene investigated, were not found to be differentially methylated[361]. *Melanin-concentrating hormone receptor 1* (*MCHR1*) plays a significant role in regulation of energy balance, and body weight and polymorphisms in this gene were previously shown to influence obesity risk. Allele specific methylation on *MCHR1* was investigated in adult blood samples, and it was evidenced that DNA methylation is negatively associated with BMI for one of the alleles examined [362]. Elsewhere, women of childbearing age with higher BMI, waist circumference, or % body fat were more likely to have lower LINE-1 methylation in peripheral blood [363]. In addition, in an investigation of methylation regions showing large inter-individual variability, methylation at four of these regions was shown to covary with BMI at two separate time points. These regions were located in or near genes *PM20D1, MMP9, PRKG1*, and *RFC5* [364].

On the other hand, one study investigating body composition measures and global genomic DNA methylation did not evidence an association between methylation and any index of adiposity such as BMI, waist circumference, subcutaneous and visceral fat, % fat mass, and fat-free mass [365]. Similarly, in a multi-ethnic birth cohort in New York City, adult BMI was not found to be associated with adult DNA methylation[317], nor was BMI found to be associated with peripheral blood global methylation among healthy controls or breast cancer cases in another investigation [171]. In another New York cohort of cancer free patients, *Interleukin-6 (IL6)* promoter methylation was also not associated with BMI [366].

In contrast to the conflicting findings on peripheral blood, overweight/obesity was found to be associated with DNA methylation in tissue biopsies. Tao et al. (2011)[346], studied promoter methylation for *E-cadherin, p16*, and *RAR-β2* in breast tumour biopsies, and even though individual gene methylation was not associated with BMI or other adiposity measures, a greater waist-to-hip ratio was associated with increased promoter methylation of 1 or more of the 3 genes in postmenopausal samples [367]. In normal autopsy kidney tissue, *RASSF1A* promoter methylation was increased with increased adiposity, and adiposity was identified as an independent predictor of *RASSF1A* methylation[368]. In another study on colon tumour biopsies, methylation status of 5 CpG island markers (*MINT1, MINT2, MINT31, p16* and *hMLH1*) was assessed to derive a CpG island methylator phenotype (CIMP). It was found that obese individuals were more than 2 times more likely to have a CIMP-low tumour than normal weight individuals [369].

Despite the indication that overweight/obesity is associated with DNA methylation, the studies above are not informative on the temporality of the association, since in these studies the change in methylation could have preceded obesity rather than be a consequence of it. The studies below investigated the temporal nature of the association, and are thus more informative, suggesting a bi-directional relationship.

Three studies investigated methylation at birth and determined that at birth methylation influences weight in children [370–372]. In one study, two differentially methylated regions on the *IGF2* gene were analysed in cord blood leukocytes and weight was assessed at age 1 year. Methylation at one of the two regions, referred to as *H19* and comprised of 4 CpG sites within a sequence motif to which zinc finger protein binds, was higher in infants with weight-for-age (WFA) in the highest 85% percentile compared to infants with a lower WFA [370]. Similarly, Godfrey et al. (2011)[371] evidenced that methylation at genes *retinoid X receptor-α* (*RXRA*) and *endothelial nitric oxide synthase* (*eNOS*) was associated with childhood fat mass at age 9 years, but replication in a second cohort only supported the first association. Lastly, 29 differentially expressed genes, between children in the highest and lowest tertiles of BMI, were analysed in cord blood samples. Five genes had methylation <1% or more than 99% and were thus not analysed further. Of the remaining 24 genes, *CDK1C* and *EPHA1* had CpG sites which demonstrated consistent and positive associations with BMI and fat mass at age 9 years. In addition, associations were observed between methylation at *CASP10* and BMI, methylation at *HLA-DOB* and NID2 and fat mass, and methylation at *MMP9* and *MPL* and lean mass at age 9 years. However, given the multiple tests in this study, larger studies are needed to validate the significance of these results. In addition to the studies investigating at birth methylation, in another prospective study, Perng et al. (2013)[373], evidenced that LINE-1 methylation in

school aged boys was not associated with BMI or skinfold thickness changes in the follow-up years, however, LINE-1 methylation significantly predicted central adiposity changes. In girls, none of the adiposity measures examined were significantly associated with LINE-1 methylation [373].

Studies on weight loss suggest that weight loss influences methylation and DNA methylation influences weight loss capacity [374–376]. Milagro et al. (2011)[374], compared high and low responders to an 8 week hypocaloric diet intervention, with respect to their basal and endpoint DNA methylation patterns. They revealed several CpGs in the *ATP10A* and *CD44* genes statistically significantly methylated at baseline between the two responder groups suggesting that these methylation differences could predict weight loss response. Also, in the high responders groups, *WNT1* promoter CpGs were found to be more methylated at endpoint suggesting that weight loss impacts on *WNT1* methylation. Perhaps more importantly, CpG methylation at *ATP10A* and *WNT1* was significantly modified as a result of weight loss[374]. In another study on obese individuals undergoing gastric by-pass surgery, global cytosine methylation in skeletal muscle was not found to differ between obese and normal weight individuals, or be altered by weight loss. Out of the 14 metabolic genes analysed, promoter methylation of 11 genes was normalised by weight loss to levels observed in healthy individuals. In addition, bisulphite sequencing determined that the promoter DNA methylation pattern of the *PGC-1α* and *PDK4* genes was restored to non-obese levels after gastric bypass associated weight loss[375]. Lastly, baseline differences in peripheral blood DNA methylation were shown to determine weight loss response in an intervention program. Out of 27,000 CpG sites examined, 5 regions in or near *AQP9, DUSP22, HIPK3, TNNT1* and *TNNI3* genes were differentially methylated between high and low-responders, while a calculated methylation score based on the most differentially methylated CpGs between the two responder groups, was also associated with changes in weight [376].

Despite the heterogeneity in study designs, exposures and outcomes, the studies above support that overweight/obesity can influence the epigenome.

Table 2 – Studies investigating the association between overweight/obesity and DNA methylation

First author, year	Subjects, N	Sample type	Exposure	Outcome	Analysis and Effect Estimate	P-value	Confounders
Slattery, 2007[369]	Colon tumours, 1154	Colon tumour tissue	BMI[∝] categories: Obese(BMI≥30kg/m^2, n=337), Overweight (BMI 25-29kg/m^2, n=742), Normal (BMI <25kg/m^2, n=674)	CIMP[∝] low phenotype: 0 or 1 methylated CpGs out of 5 examined at *MINT1, MINT2, MINT31, p16*, and *hMLH1* genes using methylation specific PCR	Logistic regression with low CIMP as the outcome and BMI categories as exposure: **Overweight vs. Normal:** OR=1.3 (1.0-1.6) **Obese vs. Normal:** OR=2.0 (1.5-2.6)	 NA[∝]	Age, sex, energy intake, activity, calcium density, smoking amount, fiber density, NSAIDs and alcohol
Peters, 2007[368]	Kidney normal autopsy tissues, 78	Normal autopsy kidney tissue	BMI categories: Adiposity (BMI≥30kg/m^2, n=17) Overweight (BMI 25-30kg/m^2, n=20) Normal (BMI 20-25kg/m^2, n=30)	*RASSF1A* promoter methylation using bisulphite sequencing	Univariate analysis between methylation % and BMI categories.	0.018	NA
Terry, 2008[317]	Adult women, 92	Peripheral Blood DNA	BMI: ≥25 vs. <25kg/m^2	Genomic DNA methylation using the [H^3]-methyl acceptance assay	Linear regression – outcome: methylation, exposure: BMI β=-0.21 (-0.48-0.05)	 NA	Prenatal smoke exposure, race, family SES, birth weight, birth length, age at menarche, childhood passive smoke exposure, smoking status, parity, age at first birth
Choi, 2009[171]	Women with incident early-stage breast cancer, n=179 and cancer free women, n=180	Peripheral Blood DNA	BMI: ≥30kg/m^2, n=104 25-29kg/m^2, n=106 <25kg/m^2, n=99	Genomic methylation measuring 5-methyldeoxycytosine (5-mdC) levels,	Linear regression of methylation with BMI categories as continuous variable: **Among all subjects:** β=0.009 **Among controls:** β=0.012	 0.717 0.730	Age and race
Wang, 2010[359]	Lean and Obese subjects aged 14-18 years, 14	Peripheral blood DNA	BMI: Obese (n=46) vs. lean (n=46) subjects	Pyrosequencing validating genome-wide methylation results	T-test on mean methylation values: ***UBASH3A*:** obese mean - 45.1% lean mean - 41.8% ***TRIM3*:** obese mean - 8.76% lean mean - 9.97%	 0.008 0.001	Age
Stepanow,	Adults, 39	Peripheral	BMI	Methylation status at	Pearson Correlation between		

2010[362]		blood DNA	(continuous variable)	MCHR1 using PCR and genotyping: Allele GT Allele AC	methylation % and BMI: r=-0.814 r=0.057	 0.024 0.897	 NA
Tao, 2011[367]	Adult women with breast cancer, 803	Breast tumour DNA	BMI: ≥30kg/m^2, n=214, 25-30kg/m^2, n=188, 20-25kg/m^2, n=164 Waist-to-hip Ratio (WHR) <0.78, n=122 0.78-0.81, n=109 0.82-0.87, n=150 ≥0.88, n=214 Lifetime weight change <4.1kg, n=120 4.1-8.0kg, n=138 8.1-12.4kg, n=164 >12.4kg, n=144	Promoter methylation at*: E-cadherin, p16*, and *RAR-ß2* using methylation specific PCR: methylation of ≥1 of 3 genes	Unconditional logistic regression: **BMI** ≥30 vs. <25kg/m^2, OR=1.02 (0.64-1.61) **WHR** ≥0.88 vs. <0.78 OR=1.85(1.10-3.11) **Weight Change** >12.4kg vs. ≤4.1kgs OR=0.91 (0.53-1.58)	 0.98 0.02 0.83	Age, race, education, age at menarche, age at first birth, age at menopause, family history of breast cancer, history of benign breast disease, energy intake, alcohol drinking status, smoking status, and oestrogen receptor status
Zhang, 2011[365]	Healthy adults, 161	Peripheral blood DNA	BMI (continuous), Waist Circumference (continuous), Total adipose tissue (continuous)	LINE1 methylation using pyrosequencing	Linear regression: **BMI**(continuous): β=0.06 (-0.37, 0.50) **Waist Circumference** (continuous): β=-0.20 (-0.64, 0.24) **Total adipose tissue** (continuous): β=-0.16 (-0.65, 0.32)	NA	Adjusted for age, gender and race/ethnicity
Godfrey, 2011[371]	Cohort 1: Children age 9years, 78 Cohort 2: Children age 9years, 239	Umbilical cord tissue DNA	Total fat mass and Percentage fat mass at age 9 years (continuous)	Umbilical cord tissue DNA methylation at 68CpG sites near genes using a mass-spectrometry based method: *retinoid X receptor-a (RXRA), endothelial nitric oxide synthase (eNOS), superoxide*	All effect estimates per SD change in CpG methylation: **Cohort 1** ***eNOS* chr7:150315553+:** 19% change in total fat mass 11% change in percentage fat mass ***RXRA* chr9:136355885+:** 23% change in total fat mass 14% change in percentage fat mass	 0.002 0.009 0.001 0.005	 *RXRA* chr9:136355885+ methylation and sex *eNOS* chr7:150315553+

				dismutase-1 (SOD1), interleukin-8 (IL8), and *phosphoinositide-3-kinase, catalytic, d-polypeptide (PI3KCD)*	**Cohort 2** ***eNOS* chr7:150315553+:** 21% change in total fat mass 13% change in percentage fat mass ***RXRA* chr9:136355885+:** 6% change in total fat mass 4% change in percentage fat mass	 0.002 0.002	methylation and sex birth weight and maternal age, smoking, and BMI
Michels, 2011[357]	Infants, 319	Cord Blood	Birthweight: High/Low (n=91) Normal (n=227)	LINE-1 methylation using pyrosequencing	Linear regression: **low vs. normal weight** β = -0,82 **high vs. normal weight** β = -0,43	 0.007 0.036	Maternal age at delivery, maternal ethnicity, maternal smoking, sex of the child, pre-term birth
Piyathilake, 2011[363]	Women of child-bearing age, 470	Peripheral Blood DNA	BMI: ≥25 vs. <25kg/m^2 Waist Circumference: >88cm vs. ≤88cm % body fat: >33% vs. ≤33%	Genomic LINE-1 methylation using pyrosequencing	Logistic regression – outcome: methylation < or > median, exposure: **BMI** OR=2.0 for lower LINE-1 methylation **Waist Circumference:** OR=1.9 for lower LINE-1 methylation **% body fat:** OR=1.8 for lower LINE-1 methylation	 0.003 0.005 0.01	
Milagro, 2011[374]	Adult men with excess body weight participating in an 8-wk energy-restricted diet treatment, 25	Peripheral blood DNA	High responders – lost >10% of initial weight (n=6) Low responders – lost <5% of initial weight (n=6)	Genome-wide methylation at ~27k CpG sites using the Illumina HumanMethylation 27 array platform followed by validation of significant sites by microarray analysis in all 25 subjects	Baseline: comparison high vs. low responders: *ATP10A* and *CD44* significantly more methylated in the responder group (separate effect estimates for each CpG) End of treatment: comparison high vs. low responders: *WT1* significant hypermethylation in the high responder group (separate effect estimates for each CpG) All subjects: Paired t-test before and after treatment: *ATP10A* CpG18 *WT1* CpG21 differentially methylated	 P<0.05 P<0.05	NA
Milagro,	Normal, over-	Peripheral	BMI:	*CLOCK* gene	T-test on mean methylation values:	NA	NA

2012[361]	weight and obese women following a 16wk weight reduction program, 60	blood DNA	Overweight+obese (n=40) vs. normal weight (n=20)	methylation *BMAL1* gene methylation *PER2* gene methylation using a mass-spectrometry based method	Significant for ***CLOCK*** CpGs 1, 5-6 and 8 Significant for ***BMAL1*** CpGs 6-7,8, 15 and 16-17		
Kuehnen, 2012[358]	Children, 107	Peripheral blood DNA	BMI: Obese (n=71) vs. Normal weight (n=36)	*POMC* exon3 CpG island using methylation specific PCR	T-test of mean methylation between obese and normal weight – exact means not available	<0.001	NA
Perkins, 2012[370]	Children age 1 year, 204	Umbilical cord blood leukocytes	Weight for Age (WFA) >85th per-centile vs. WFA ≤85th percentile	Umbilical cord methylation at two *IGF2* differentially methylated regions using pyrosequencing: 3 CpGs at the IGF2 DMR and 4 CpGs at the H19 DMR	Generalized regression model: WFA >85th percentile vs. WFA ≤85th percentile: **H19DMR** mean methylation fraction ≥75th percentile: OR=3.12 (1.13-8.60) **IGF2DMR** mean methylation fraction ≤25th percentile: OR=1.19 (0.36-3.39)	0.03 0.98	maternal cigarette smoking, breastfeeding, race, and postnatal caloric intake.
Relton, 2012[372]	Children age 9years, 178	Cord blood DNA	BMI and fat mass at age 9 years (continuous)	Cord blood methylation of 29 genes previously identified as differentially expressed according to BMI using the GoldenGate Assay Kit	All effect estimates per 1% increase in methylation: ***CDKN1C:*** 2.08% increase in BMI 5.16% increase in fat mass ***EPHA1:*** 0.80% increase in BMI 1.84% increase in fat mass ***CASP10_P2*:** 2.13% decrease in BMI ***HLA_DOB3*:** 0.90% decrease in fat mass ***NID1_P*:** 1.19% decrease in fat mass	 0.031 0.037 0.048 0.036 0.02 0.039 0.035	 age, sex, batch age, sex, batch, height age, sex, batch age, sex, batch, height age, sex, batch age, sex, batch, height age, sex, batch, height

Zhang, 2012[366]	Adult cancer free subjects, 165	Peripheral Blood DNA		Genomic (LINE-1) and gene-specific (IL-6) methylation using pyrosequencing	No association with BMI		
Huang, 2012[360]	Adolescents age 17 years, 315	Peripheral blood DNA	Skinfold thickness, subcutaneous and visceral adiposity, weight (continuous)	12 CpG sites within the *IGF2*/H19 imprinting control region (ICR) using a mass-spectrometry based method	~5% increase in mean IGF2/H19 ICR DNA methylation: **Skinfold thickness** 0.24% standard deviations increase (2.5 mm in boys and 2.0 mm in girls) **Subcutaneous adiposity** 0.18% standard deviations increase (1.9 mm in boys and 2.0mm in girls)	 <0.005 <0.05	NA
Perng, 2013[373]	School-age children, 335 (boys, n=254; girls, n=299)	Peripheral blood DNA	BMI-for-age Waist Circumference Skinfold thickness	Genomic LINE-1 methylation using pyrosequencing (Quartile 4 vs. Quartile 1)	Linear regression with the anthropometric variable as the outcome and LINE-1 as exposure. All effect estimates are Z-score changes/year in the outcome comparing Q4 vs. Q1 methylation: <u>Boys:</u> **BMI-for-age:** -0.06 (-0.12, 0.01) **Waist Circumference:** -0.09 (-0.17, -0.02) **Skinfold thickness:** -0.04 (-0.12, 0.03) <u>Girls:</u> **BMI-for-age:** -0.01 (-0.07, 0.06) **Waist Circumference:** -0.01 (-0.08, 0.06) **Skinfold thickness:** 0.00 (-0.07, 0.07)	P-trend between Quartiles 0.21 0.02 0.28 0.65 0.99 0.85	Age and socioeconomic status
Barres, 2013[375]	Obese women undergoing gastric by-pass surgery (RYGB) (n=8) and normal weight women (n=9)	Skeletal muscle biopsies	Weight loss after RYGB	Methylation of 14 genes using bisulphite sequencing: *PDHA1, MYH2, PPP1R3C, PTP4A3M, PDP1, SLC16A1, PRKAA1, CKMT1B, EXOC5, ACOX1,*	RYGB surgery weight loss restored the methylation of the following 11 genes to normal-weight levels (exact values not given): *PDHA1, MYH2, PPP1R3C, PTP4A3M, PDP1, SLC16A1, PRKAA1, CKMT1B, EXOC5, ACOX1, ACACB*	NA	Age

				ACACB, INPP5J, MRPL41, OXA1L DNA methylation pattern of the *PGC-1a* and *PDK4* promoters on proximal region relative to the transcription start site	RYGB sur- gery weight loss decreased *PGC-1* a promoter methylation and increased PDK4 promoter methylation to non-obese levels		
Moleres, 2013[376]	Overweight or obese adolescents undergoing a 10wk weight loss intervention, 107	Peripheral blood DNA	High responders: losing >1.1 BMI-standard deviation score (BMI-SDS) (n=12) Low responders: Losing <0.4 BMI-SDS after treatment (n= 12)	Genome-wide methylation at ~27k CpG sites using the Illumina HumanMethylation 27 array platform followed by validation by microarray analysis in all 107 subjects of significant sites corresponding to *AQP9, DUSP22, HIPK3,TNNI3, TNNT1, F12*, and *PTPRG* genes	Baseline: comparison high vs. low responders: *TNNT1, AQP9, TNNI3, HIPK3, and DUSP22* differentially methylated (separate effect estimates for each CpG) Differences in basal DNA methylation of several CpGs in *AQP9, DUSP22, HIPK3, TNNI3*, and *TNNT1* genes were significantly associated with changes in body weight, BMI-SDS, waist girth, and body fat mass after the intervention (separate effect estimates for each CpG)	NA	NA

∝ BMI, body mass index; CIMP, CpG island methylator phenotype; NA, not available

1.4.4 Age at Menarche/Menopause and DNA methylation

As above, a literature review of available published studies to May 1st, 2013 was conducted to identify studies investigating the association between age at menarche and/or age at menopause and DNA methylation. However, the literature on the relationship between these reproductive factors and DNA methylation is very limited and the evidence is controversial (Table 3).

Terry et al. (2008)[317] investigated factors that impact on breast cancer risk and evidenced that a greater birth height, a later age at menarche, nulliparity, and a later age at first birth were associated with higher global DNA methylation levels. On the contrary, in a study on healthy controls and breast cancer cases, age at menarche, age at first birth, parity, menopausal status and age at menopause were not significantly associated with peripheral blood global methylation [171]. In another study, published as an abstract, age at menarche was not associated with differential methylation at any of the 27,578 methylation sites examined, but 1,452 CpGs were differentially methylated in postmenopausal women compared to premenopausal women [377].

Therefore, of the secular trends evaluated, age at menarche and age at menopause provide the least evidence regarding an association with DNA methylation and more studies are needed to understand whether or not age at menarche and/or age at menopause influence DNA methylation.

Table 3 – Studies investigating the association between age at menarche and/or age at menopause and DNA methylation

First author, year	Subjects, N	Sample type	Exposure	Outcome	Analysis and Effect Estimate	P-value	Confounders
Terry, 2008[317]	Adult women, 92	Peripheral Blood DNA	Age at menarche (continuous)	Genomic DNA methylation using the [H^3]-methyl acceptance assay	Linear regression – outcome: methylation, exposure: age at menarche (continuous) β=-0.07 (-0.13 - -0.004)	 NA[∝]	Prenatal smoke exposure, race, family SES, birth weight, birth length, age at menarche, childhood passive smoke exposure, smoking status, parity, age at first birth
Choi, 2009[171]	Women with incident early-stage breast cancer, n=179 and cancer free women, n=180	Peripheral Blood DNA	Age at Menarche: <13, n=160 ≥13, n=155 Age at Menopause: <50, n=91 ≥50, n=90	Genomic methylation measuring 5-methyldeoxycytosine (5-mdC) levels	Linear regression of methylation with exposure as continuous variable: **Age at menarche** **Among all subjects:** β=-0.036 **Among controls:** β=-0.055 **Age at menopause** **Among all subjects:** β=-0.026 **Among controls:** β=-0.042	 0.375 0.334 0.271 0.220	Age and race
Nichols, 2013[377]	Women in the Sister Study, n=162	Peripheral Blood DNA	Menopausal Status Age at menarche	Genome-wide methylation at ~27k CpG sites using the Illumina HumanMethylation 27 array platform	Post-menopausal vs. Pre-menopausal women: 1452 CpG sites residing in 1220 CG islands were differentially methylated between the two groups No differentially methylated CpGs according to age at menarche	 FDR: <0.05	NA

[∝] NA, not available

1.4.5 Height and DNA methylation

An extensive and systematic literature search for publications investigating the association between height and DNA methylation yielded only three studies, only one of which evidences a significant association (Table 4). In a recent study, investigating 29 differentially expressed genes between children in the highest and lowest tertiles of BMI, cord blood DNA methylation at *alkaline phosphatase* (*ALPL*) *and Interferon regulatory factor 5* (*IRF5*) genes was associated with height at age 9 years. The associations remained even after adjustment for age, sex, batch, fat mass, and lean mass, and the association between methylation at *ALPL* and decreased height withstood correction for multiple testing [372].

However, among 17-year-olds, methylation at the *IGF2/H19* imprinted region in peripheral blood was not shown to be significantly and positively correlated with height[360]. Similarly, in a prospective study among school aged children, LINE-1 methylation did not significantly predict height changes during follow-up [373].

Therefore, whether height influences DNA methylation cannot be safely concluded based upon the already available literature.

Table 4 – Studies investigating the association between height and DNA methylation

First author, year	Subjects, N	Sample type	Exposure	Outcome	Analysis and Effect Estimate	P-value	Confounders
Relton, 2012[372]	Children age 9years, 178	Cord blood DNA	Height at age 9 years (continuous)	Cord blood methylation of 29 genes previously identified as differentially expressed according to BMI[∝] using the GoldenGate Assay Kit	All effect estimates per 1% increase in methylation: ***ALPL_P*:** 0.15% decrease in height ***IRF5_P*:** 0.42% decrease in height ***IRF5_E*:** 0.29% decrease in height	 0.0002 0.022 0.026	 age, sex, batch age, sex, batch age, sex, batch
Huang, 2012[360]	Adolescents age 17 years, 315	Peripheral blood DNA	Height (continuous)	12 CpG sites within the *IGF2*/H19 imprinting control region (ICR) using a mass-spectrometry based method	~5% increase in mean *IGF2*/H19 ICR DNA methylation: **Height** 0.05% standard deviations decrease (0.365 cm in boys and 0.325 cm in girls)	 >0.05	 NA[∝]
Perng, 2013[373]	School-aged children, 335 (boys, n=254; girls, n=299)	Peripheral blood DNA	Height-for-age	Genomic LINE-1 methylation using pyrosequencing (Quartile 4 vs. Quartile 1)	Linear regression with height as the outcome and LINE-1 as exposure. Effect estimate is Z-score change/year in height comparing Q4 vs. Q1 methylation: <u>Boys:</u> **Height-for-age:** 0.02 (-0.03, 0.08) <u>Girls:</u> **Height-for-age:** 0.02 (-0.04, 0.08)	 0.55 0.54	Age and socioeconomic status

[∝] BMI, body mass index; NA, not available

1.5 Hypothesis and Study Aims

Despite the limited number of available studies for some of the exposures examined, all exposures were shown to be associated with methylation in at least 1 publication (Table 5).

Table 5 – Exposures undergoing secular trends and DNA methylation.

Number of studies investigating the association between the exposures of interest and DNA methylation and the number of these studies that have evidenced at least one statistically significant association between exposure and DNA methylation

Exposure	Studies investigating the association between exposure and DNA methylation	Studies evidencing at least one significant association between exposure and DNA methylation
Early life SES	8	6
Adult life SES	4	3
Nutrition	47	34
Overweight/Obesity	21	16
Age at Menarche	3	1
Age at Menopause	2	1
Height	3	1

Based on the discussed literature, it is hypothesised that factors undergoing secular changes, and more specifically SES, nutrition, overweight/obesity, height and age at menarche as well as age at menopause leave their mark on the epigenome, by influencing DNA methylation. DNA methylation is one way through which these "exposures" can become biologically embedded and influence disease risk, and especially breast cancer risk. As these factors continue to display secular trends, especially in low-middle income countries, such an investigation is imperative so as to be able to better understand how and at what degree these secular trends will influence disease prevalence.

A first descriptive approach of this hypothesis is carried out using global DNA methylation measures, using the LUminometric Methylation Assay, on 396 individuals and using whole genome methylation data, covering more than 450,000 CpGs in 670 subjects free of disease at time of blood collection.

2 Subjects and Methods

The present investigation included both global methylation and whole genome methylation analyses.

2.1 Subjects and Methods for Global Methylation Analyses

2.1.1 Study Participants

Epidemiologic data and blood samples collected from the EPIC study were used for global methylation analyses. EPIC is an on-going study designed to investigate diet, nutrition, lifestyle and environmental factors with respect to cancer incidence. The cohort consists of 519,978 participants from 23 centres in 10 European countries - Denmark, France, Germany, Greece, Italy, the Netherlands, Norway, Spain, Sweden and the United Kingdom. Information on lifestyle, diet, anthropometric measures and environmental exposures was collected using questionnaires at recruitment and variables were standardised across the different participating centres. Blood was also collected from the majority of subjects at recruitment [378]. For this investigation, 300 female breast cancer cases and 300 female matched controls from the EPIC cohort were chosen. Of these, 77 subjects were initially excluded: 1 subject was a duplicate, there was not enough DNA for 24 subjects, and 52 samples produced no or a weak signal. Of the remaining 523 subjects with reliable measurements, 376 women were investigated in this specific study, who remained free of cancer for at least 5 years following blood collection. All participants signed an informed consent and the study protocol was approved by the ethical review boards of the International Agency for Research on Cancer and by the local committees of the participating centres.

2.1.2 Methods for Global Methylation Measurement

LUminometric Methylation Assay (LUMA) was used to quantify global methylation levels [379,380] in Peripheral Blood Leucocytes (PBLs) of subjects whose blood was collected at recruitment. Genomic DNA was extracted using standard protocols. Genomic DNA was digested with *Hpa*II+*Eco*RI or *Msp*I+*Eco*RI restriction enzymes in 2 separate reactions, set up in a 96-well plate format. Polymerase extension assay by Pyrosequencing was employed to determine cleavage. The LUMA method was validated using DNA controls of known DNA methylation status as described in [381]. In the assay, 5-Aza-dC treated and CpG methylated Jurkat genomic DNA (New England Biolabs, Ipswich, MA) were used as methylated and unmethylated control samples, respectively. A measure of % global methylation was calculated using the ratio of DNA cleavage by methylation sensitive (*Hpa*II) and methylation insensitive (*Msp*I) restriction enzymes based on the equation:

Genome wide Methylation (%) = [1-(*Hpa*II/*Eco*RI)/(*Msp*I/*Eco*RI)*100

2.1.2.1 LUminometric Methylation Assay

LUMA is a cytosine extension assay where DNA is digested by methylation sensitive restriction endonucleases (*Hpa*II). These enzymes cut DNA at CpG sites that have lost their methylation leaving 5′-CG overhangs. Cleavage is followed by single nucleotide extension using either radiolabelled [^{3}H]dCTP or biotinylated dCTP. Therefore incorporation of the radiolabelled dinucleotides is directly related to the number of digested sites and thus inversely associated with DNA methylation. In another reaction, DNA is digested by methylation insensitive restriction endonucleases (*Msp*I), where DNA is cleaved at CpG sites, irrespective of DNA methylation. In addition to *Hpa*II or *Msp*I, *Eco*RI is added to both reactions and *Eco*RI cleavage acts as a normalisation reference for the amount of DNA. *Eco*RI cleavage produces 5′-AATT overhangs which are filled with dNTPs in a polymerase extension assay. The addition of nucleotides triggers a cascade reaction involving ATP which creates a proportional amount of light[379,380].

In LUMA, the results are presented in a luminometric output with different light peaks produced for each nucleotide added in a sequential manner. The ratio of light intensity when dCTPs are added to the *Hpa*II digested DNA (proportional to *Hpa*II cleavage) to the light intensity when dCTPs are added to the *Msp*I digested DNA (proportional to *Msp*I cleavage) is used to determine % global methylation. The light intensities when dATPαS and dTTP are added are used as a normalisation method as these nucleotides are incorporated to *Eco*RI cleavage sites and their light intensities (proportional to *Eco*RI cleavage) should be equal. Examples of the LUMA luminometric output are presented in Figure 15.

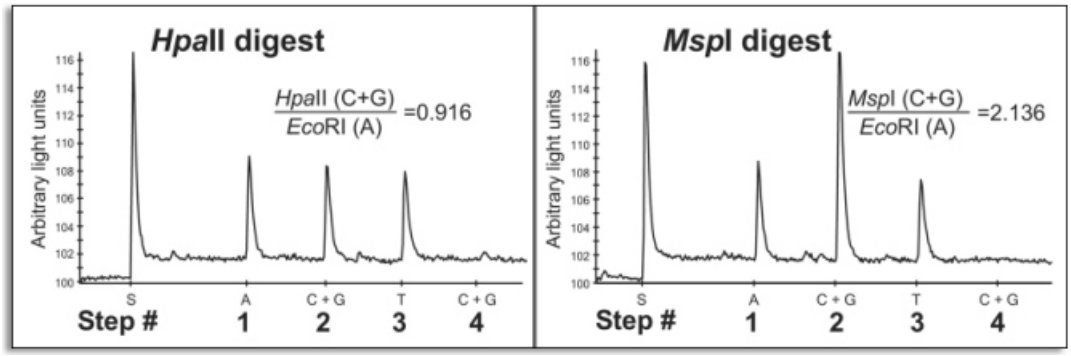


Figure 15 – Luminometric output

HpaII and MspI digestion with EcoRI followed by LUMA analysis in a Pyrosequencer. The graphs show the luminometric output from two representative LUMA runs of human lymphocyte DNA, using HpaII + EcoRI, and MspI + EcoRI. The A and T peaks denote the additions of dATPαS and dTTP, respectively, the nucleotides of which are substrates for filling in all EcoRI generated T and A overhangs. Reproduced with permission from [380].

2.1.3 Statistical Analyses

The distribution of the variables under investigation across quartiles or medians of percent global methylation was first explored. The quartile cut offs for global methylation were the 25th, 50th, and 75th percentile global methylation values in controls (subjects who remained free of cancer throughout the study's follow-up period). For continuous variables, the non-parametric equivalent of one-way analysis of variance (ANOVA), the Kruskal-Wallis test, was used as genome-wide methylation was not normally distributed. For categorical variables, chi-square test was performed.

Further investigation of the variables under study was performed using logistic models because the resulting profile of global methylation distribution was skewed and several commonly used transformations failed to normalise it. In addition, various Generalized Linear Models (GLMs) investigated failed to adequately describe the outcome distribution. In order to perform unconditional logistic regression, the methylation outcome was dichotomised to above and below median methylation. For the purposes of simplicity, for this analysis, below median methylation will be referred to as "hypomethylation" whereas above median methylation will be referred to as "hypermethylation". Investigated variables were included in separate regression models as continuous and categorical exposures when relevant, as described in more detail in the following sections. Based on the available literature, all logistic regression models were adjusted for centre, plate number, case-control status, age at blood collection and smoking status. Even though all women were healthy at blood collection, and even though all diagnoses within 5 years of blood collection were excluded, adjusting for case/control status was thought as imperative given the unknown latency periods of the different forms of breast cancer. Depending on the variable under investigation, additional confounders were considered. All confounders were entered into the model as continuous variables with the exception of smoking status which was treated as categorical – past, never, present. All subjects with LUMA measurements were women, therefore adjustment for sex was not necessary. For each logistic regression model built, regression diagnostics were performed to examine multi-collinearity (*collin*), significance of the model (*linktest*), and goodness of fit (*lrtest, fitstat*).

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

2.1.3.1 Socioeconomic Status and Global Methylation

The available variable that could be used as a proxy for the investigation of **adult socioeconomic status** and global methylation was highest level of education attained.

Highest level of education attained was treated as a categorical variable in the logistic regression model: none (reference), primary school, technical/professional school, secondary school, and university. Given the associations between education and weight as well as dietary choices, and given the associations of these

variables with DNA methylation in the literature, models investigating SES were further adjusted for BMI, and daily alcohol and folate intakes.

As a proxy for **early life socio-economic conditions**, the year of birth of subjects was investigated. Individuals born during World War II (WWII), 1939-1944, were assumed to have been exposed to stressful conditions of less than optimal hygiene and nutrition availability. Whether birth during the war is associated with adult global DNA methylation was investigated adjusting for the baseline covariates.

2.1.3.2 Nutrition and Global Methylation

Several aspects of nutrition were investigated for their influence on global DNA methylation.

At first, total caloric intake as well as each food group and each nutrient variable was independently assessed for its influence on DNA methylation, by comparing the distribution of the variable across the two methylation groups (hypo- and hyper-methylation). **Folate** and **alcohol**, because of their key roles in the one-carbon metabolism which influences methylation, as well as **total caloric intake** were also investigated in unconditional logistic regression models. All three variables were treated as categorical (quartiles of intake) since unit increases are too small to have any biological effect. All models were adjusted for centre, plate number, case-control status, age at blood collection and smoking status as well as for education, BMI and physical activity as these factors have been shown to influence both diet and methylation. The models with alcohol and folate intake were further adjusted for total caloric intake, and the model on total caloric intake was also adjusted for folate and alcohol intakes. Lastly, an interaction term between alcohol and folate was tested in the same model used to investigate alcohol and/or folate.

Next,food group intakes were used to calculate each person's adherence to the **Mediterranean diet pattern**. Adherence was assessed using the Trichopoulou score described elsewhere [91]. Briefly, median intakes of the following food groups were calculated for all subjects: vegetables, legumes, fruit/nuts/seeds, cereals, fish, polyunsaturated to saturated fatty acid ratio, dairy, meat, and alcohol. For the beneficial components of the Mediterranean diet, namely vegetables, legumes, fruit/nuts/seeds, cereals, fish, and polyunsaturated to saturated fatty acid ratio, a subject received a score of +1 if their intake was above median. For the detrimental components, namely dairy, meat, and alcohol a subject received a score of 0 if their intake was above median. In contrast to previous publications, where moderate consumption of alcohol (5-25g/day) received a score of +1, here, because any alcohol consumption is detrimental to cancer risk, all women with below median consumption (2.69g/day) received a score of +1. Adherence to the Mediterranean diet was assessed for its impact on DNA methylation by treating the adherence score as a continuous and as a categorical variable in separate logistic regression models again adjusting for centre, plate number, case-control status, age at blood collection and smoking status as well as for education, total caloric intake, BMI and physical activity. The score categories analysed were 0-3: low, 4-6: moderate, and 7-9: high adherence to the Mediterranean Dietary pattern.

In addition, a **methyl intake score** was calculated for each subject. Intakes of those nutrients with key roles in the one-carbon metabolism were used to obtain a total methyl intake score. It is understood that the most important dietary sources of methyl groups, second to folate, are choline and methionine. Unfortunately the last two intake values were not available. Therefore the score was based on intakes of folate, alcohol and the B vitamins, most of which are thought of as epigenetically active because they have key roles in the metabolism of carbon units. As presented in Table 6, Vitamin B1, B2, B6, B12 and folate intakes above median received a score of +1, whereas alcohol intake above median received a score of 0 due to its negative effect on the bioavailability of methyl groups. The methyl intake dietary score was assessed for its impact on DNA methylation by treating the score as a continuous and as a categorical variable in separate logistic regression models again adjusting for centre, plate number, case-control status, age at blood collection and smoking status as well as for education, total caloric intake, BMI and physical activity. The score categories analysed were 0: no, 1-2: low, 3-4: moderate, and 5-6: high methyl intake.

Table 6 – Calculation of a methyl intake dietary score

Median derived from the distribution among the 376 subjects

Nutrient	**Score if intake is above the median***
Thiamin – Vitamin B1 (mg/day)	+1
Riboflavin – Vitamin B2 (mg/day)	+1
Vitamin B6 (mg/day)	+1
Cobalamin – Vitamin B12 (µg/day)	+1
Folate (µg/day)	+1
Alcohol (g/day)	0
Total Score:	0-6

2.1.3.3 Anthropometric measurements and Global methylation

Overweight/obesity and **height** were investigated together for their effect on global methylation due to the dependence of weight on height. At first, both height and weight variables were entered together into a logistic regression model, adjusting for centre, plate number, case-control status, age at blood collection and smoking status. For this model, further adjustment for total caloric energy intake, percent energy intake from fat and sex-adjusted total physical activity was essential given their known effect on height and weight as well as their suspected influence on DNA methylation. Both height and weight were treated as categorical variables (in quartiles) because unit changes in either variable are not likely to have an effect on global methylation.

Next, given the significant correlation between height and weight, two separate models were built. One model included BMI, a measure of overweight/obesity that accounts for height's effect on weight, and the other model

included height. In the "bmi-model" BMI was treated as a categorical variable (normal: ≤25kg/m^2, overweight: >25 and ≤30 kg/m^2, obese: >30kg/m^2) and was adjusted as previously described for centre, plate number, case-control status, age at blood collection, smoking status, total caloric energy intake, percent energy intake from fat and sex-adjusted total physical activity. The "height-model" was only adjusted for centre, plate number, case-control status, age at blood collection, smoking status and total caloric energy intake since physical activity and adult fat intake do not impact on height. Height was divided into quartiles and entered into the model as a categorical exposure, since a unit change (1 cm) is not likely to manifest a biological response (change in methylation).

Following, other measures of adiposity such as **waist-to-hip ratio** and **waist circumference** were investigated for their impact on DNA methylation. Adjusting for the same confounders as in the "bmi-model", both variables were investigated as categorical (in quartiles), again because a unit change would be too small to be biologically significant.

2.1.3.4 Age at Menarche, Menopausal Status, Age at Menopause and Global methylation

Age at menarche and **age at menopause** were investigated for their influence on global DNA methylation in logistic regression models adjusting for centre, plate number, case-control status, age at blood collection and smoking status. In addition, given height's and weight's influence on age at menarche, and given that pubertal height and weight are often correlated with adult body and weight, the model on age at menarche was also adjusted for height and BMI. Both reproductive variables were treated first as continuous and then as categorical outcomes. The categories selected for age at menarche were ≤11 years, 12-14 years, and ≥15 years to signify early, normal, and late age at menarche. Age at menopause was categorised to above and below the median menopausal age which was 50 years of age. **Menopausal status** was also investigated adjusting for baseline covariates, similar to the model investigating age at menopause.

Given that much research points to the fact that the biological effect of the ages of menarche and menopause depends on the cumulative exposure to ovarian cycles throughout life [100], a new variable was computed to investigate cumulative oestrogen exposure at the point of blood collection. The variable was named **Estimated Lifetime oEstrogen Exposure (ELEE)** and crudely reflects the length of a woman's reproductive period in years. For post-menopausal women, ELEE was calculated by subtracting age at menarche from age at woman's last period. For pre-menopausal women, ELEE up to the point of blood collection was calculated by subtracting age at menarche from age at blood collection. This variable was investigated both as categorical and as continuous, adjusting for the same confounders as in the models for age at menarche.

2.2 Subjects and Methods for Genome Wide Methylation Analyses

2.2.1 Study Participants:

For the genome-wide methylation analyses, subjects from the EPIC Italy sub-cohort were used. The EPIC Italy sub-cohort consists of 47,745 subjects - 32,577 female and 15,168 male – recruited from 5 different centres – Varese, Turin, Florence, Naples, and Ragusa. From this sub-cohort 670 subjects, 335 cancer cases and 335 controls, matched on date of birth (± 5 years), seasonality of blood draw, and date of recruitment, were selected. The 670 subjects consisted of 166 breast cancer cases and their 166 matched controls, as well as of 169 colon cancer cases and their 169 matched controls.

Since the goal of this thesis was to investigate factors undergoing secular changes, we have included all 670 subjects who were healthy at recruitment, irrespective of their cancer status in the follow up, and we have included the cancer status (yes/no) as a covariate in each analysis.

However, given the long latency period of neoplasia, for each exposure, a second analysis was carried out, where only subjects who remained healthy for at least 5 years following recruitment and blood collection (n=504) were analysed.

2.2.2 Methods for Genome wide Methylation measurement

DNA was extracted from buffy coats or blood cell fractions using the QIAsymphony DNA Midi Kit (Qiagen, Crawley, UK). 500 ng of DNA was bisulphite-converted with the EZ-96 DNA Methylation-Gold™ Kit, used according to the manufacturer's protocol (Zymo Research, Orange, CA, USA). Next, the 450K DNA methylation array by Illumina (Infinium HumanMethylation 450 BeadChip) was performed on 4 µl of bisulphite-converted DNA, following the Illumina Infinium HD Methylation protocol. Briefly, a whole genome amplification step was followed by enzymatic end-point fragmentation and hybridization to HumanMethylation 450 BeadChips at 48°C for 17 h, followed by single nucleotide extension. The incorporated nucleotides were labelled with biotin (ddCTP and ddGTP) and 2,4-dinitrophenol (DNP) (ddATP and ddTTP). After the extension step and staining, the BeadChip was washed and scanned using the Illumina HiScan SQ scanner. The intensities of the images were extracted using the GenomeStudio (v.2011.1) Methylation module (1.9.0) software, which normalizes within-array data using different internal controls that are present on the HumanMethylation 450 BeadChip and internal background probes. Information on bisulphite batch, DNA amplification and fragmentation batch, as well as hybridization and extraction batch was recorded for each sample. Replicates of some DNA samples across different combinations of the above mentioned batches were used to test the quality of the method and to ensure that these processes did not significantly alter the methylation results. In addition, control beads present on the chip were used to evaluate the quality of the bisuphide conversion for each sample. Lastly, presence of any signal was used as a marker that whole genome amplification was successful.

2.2.2.1 HumanMethylation 450 BeadChip

The Illumina Infinium technology was first utilised on the Infinium HumanMethylation27 (27k) BeadChip and more recently the coverage of Illumina's array was increased to cover 485,577 CpG sites on the Infinium HumanMethylation450 (450k) BeadChip. On the 450k BeadChip emphasis was placed on gene (99% coverage) and CpG Island regions (96% coverage).

In order to be used on this chip, DNA needs to undergo bisulphite treatment to convert unmethylated cytosines to uracil. If previously methylated, cytosines remain unconverted. The DNA is then amplified.

The 450K BeadChip contains probes with two different designs – Infinium I (also called type I) and Infinium II (also called type II). Infinium I design interrogates mostly CpGs in CpG dense regions and consists of 2 probes per CpG locus – one "methylated" and one "unmethylated" probe. The 3' terminus of the "methylated" probe is designed to match the unconverted cytosine whereas the 3' terminus of the "unmethylated" probe is designed to match the converted thymine. Each probe is 50 bp long. CpG sites in this 50bp region – other than the target CpG – were assumed to have the same methylation status as the target CpG[382]. The assumption that methylation would be regionally correlated was supported by other studies that found ~90% of CpG sites within 50 bases to have the same methylation status[383]. Infinium I probes can include up to 8 CpG sites other than the target locus.

Infinium II design consists of a single probe per interrogated locus. Infinium II mostly interrogates CpG sites in areas with low CpG density. The 3' terminus of the probe compliments the base directly upstream of the query site and the addition of a labelled G base (for methylated/unconverted) or a labelled A base (for converted unmethylated cytosines to uracil) signifies the methylation status. Probes are also 50 bp long. Infinium II probes can have up to a maximum of three non-target CpG sites. However, on the probe, these non-target sites are represented by a "degenerate" R-base allowing multiple combinations of oligos to attach to the bead. Thus no assumptions are made regarding the methylation status of neighbouring CpGs[382]. The two probe designs are shown in Figure 16.

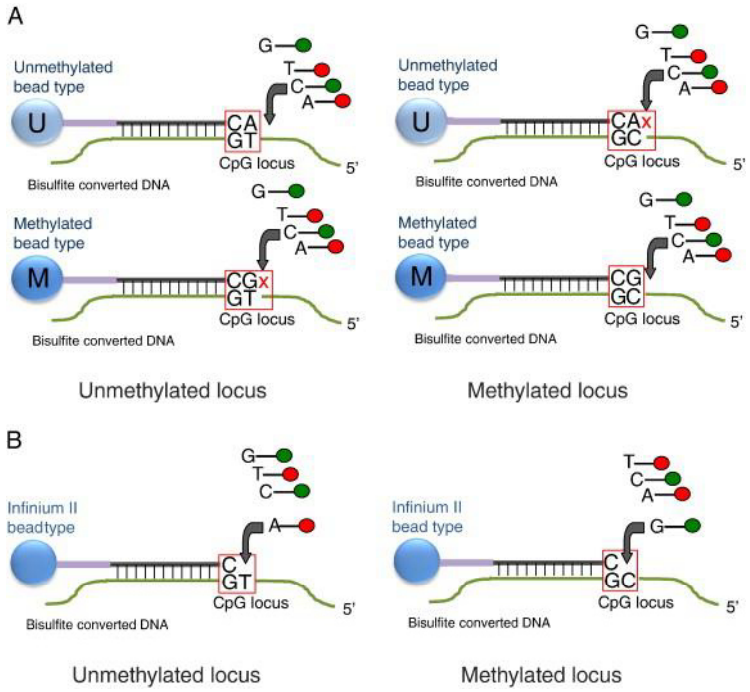


Figure 16 – Infinium Methylation Assay scheme.

1A. Infinium I assay. Two bead types correspond to each CpG locus: one bead type — to methylated (C), another bead type — to unmethylated (T) state of the CpG site. Probe design assumes same methylation status for adjacent CpG sites. Both bead types for the same CpG locus will incorporate the same type of labelled nucleotide, determined by the base preceding the interrogated "C" in the CpG locus, and therefore will be detected in the same colour channel.1B. Infinium II assay. One bead type corresponds to each CpG locus. Probe can contain up to 3 underlying CpG sites, with degenerate R base corresponding to C in the CpG position. Methylation state is detected by single-base extension. Each locus will be detected in two colours. In the current version of the Infinium II methylation assay design, labelled "A" is always incorporated at unmethylated query site ("T"), and "G" is incorporated at methylated query site ("C"). Reproduced with permission from [382].

The HumanMethylation450 (450k) BeadChip array includes 485,577 cytosine positions in the human genome: 482,421 (99.4%) are part of a CpG, 3091 are non-CpG sites and 65 are random SNPs. Hereafter the term CpG will be used to refer to all of these, unless otherwise specified.

The array covers 21,231 out of 21,474 UCSC RefGenes (98.9%) with an average of 17.2 probes per gene region. Transcripts of the RefGenes are also included. Promoter regions are divided into two different categories – the TSS200 which covers 200bp and the TSS1500 which covers 1500 bp upstream of the transcription start site. The 5' and 3' un-translated regions, the first exon and the gene bodies are also targeted [382].

Lastly, ninety six percent of CpG islands from the UCSC database (26,658 CpG islands) are covered by the array with an average of 5.63 sites per island. In addition, upstream and downstream regions immediately outside the CpG islands are also covered and these regions are called north and south shores respectively. Similarly, 2 kb regions upstream (north) and downstream (south) of the CpG islands – called shelves – are also targeted by the array.

2.2.3 Statistical Analyses

2.2.3.1 Pre-processing

2.2.3.1.1 Quality Control

Methylated and unmethylated intensities for each probe were provided by GenomeStudio software. In addition to the methylation value for each probe, corresponding detection P-values were also provided by GenomeStudio software (Illumina). The detection values indicate the confidence that can be placed on a methylation reading value. As a first step, all readings with a P-value>0.05 were considered as non-detected and replaced with 'NA'. Probes that were not detected in more than 20% of all samples and samples with <95% detected probes were excluded so as to not influence downstream pre-processing and analyses. The remaining 'NA' readings were imputed using KNN (k-nearest neighbours) method once, implemented in knn.impute function from R-CRAN, because several analyses and normalization methods cannot handle missing data.

Background noise correction was then performed as background fluorescence can contribute an additive error to each signal intensity leading to a reduced dynamic range for the methylation reading. Background correction can be performed using a simple background subtraction, where the intensities of the control probes are subtracted from the methylated and unmethylated probe intensities. However, given that signal intensities can be red or green, and given the technical variation in fluorescent signal depending on the intensity colour, dye bias also had to be taken into account using a method similar to the one described by Triche et al.[384]. This method equalises the intensities in the green and red channels to the average intensity across the two colours as measured by normalisation control probes present on the BeadChip. The analysis of other classical quality control measures (such as staining, extension, hybridization, or bisulphite conversion) provided by GenomeStudio did not reveal any major quality issues. These steps were performed by the bioinformatician of the group, Gianluca Campanella.

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

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

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

2.2.3.1.2 Normalisation

Various normalisation methods have been proposed as essential pre-processing steps for the Infinium HumanMethylation 450 BeadChip. Whole genome methylation data need to be pre-processed in order to correct for and account for a) within and between sample technical variation, b) the different design of type I and type II probes, and c) batch effects.

It is known that non-biological factors can contribute to the variability of high throughput data, such as whole genome methylation data. These non-biological factors can include variation in sample preparation, array production, and array processing [385]. However, in order to compare samples between arrays and delineate true biological variation in contrast to technical variation, differences of non-biological origin should be minimised, and most researchers agree that this is achieved through normalisation methods. The most widely used method of normalisation, which was originally developed for expression microarray studies, is quantile normalisation (QN). QN can be applied on signal intensities or on β-values and assumes that there is an underlying common distribution of intensities/methylation across chips and makes the distribution of probe intensities/methylation the same for every chip. The normalisation distribution is the distribution formed by averaging probe intensities/methylation at each quantile across chips [385].

In addition to random technical variation, differential design for type I and type II probes on the Infinium HumanMethylation 450 BeadChip calls for further correction. This is because Dedeuwaerder et al. [386] show that the two probe types give a different distribution of β-values, and that type II probes give β-values that are less accurate and less reproducible.

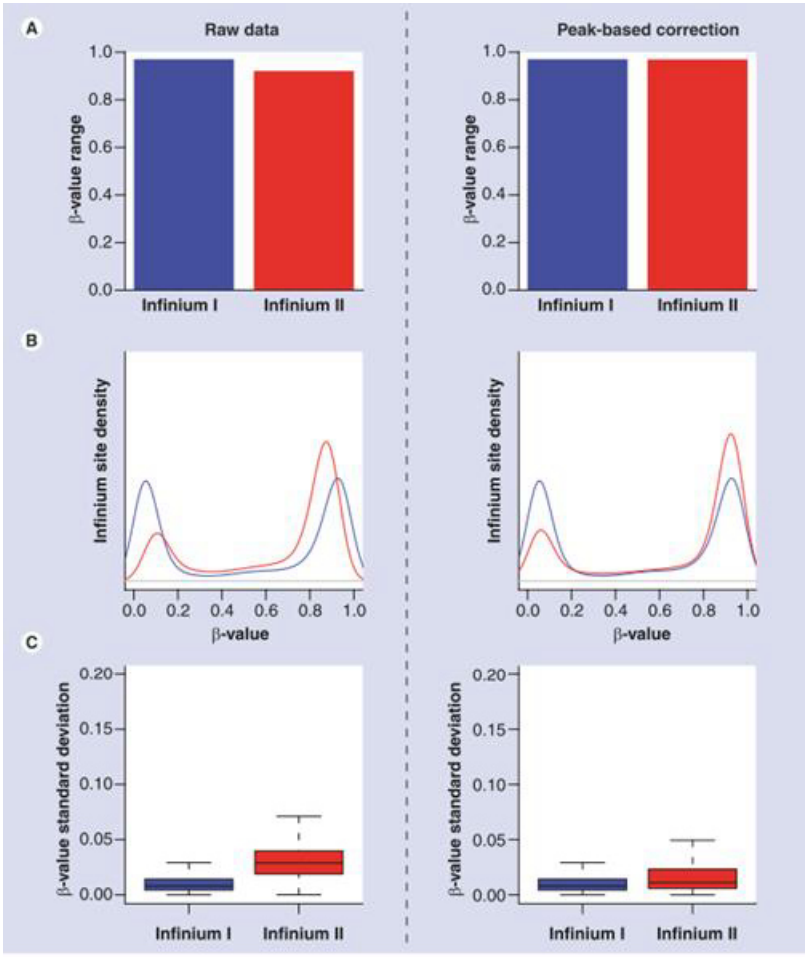


Figure 17 – Peak-based correction method for improving Infinium II (type II) data which are less accurate and reproducible than Infinium I (type I) data.

(A) Bar plots indicating the range of b-values generated for a given sample with the Infinium I and Infinium II assays. (B) Density plots of the b-values for the two Infinium assay types considered (blue: Infinium I; red: Infinium II) for a given sample. (C) Box plots of probe-wise variance between the three replicates of a given sample for the Infinium I and Infinium II probes (outliers not drawn). On the left part of the figure, b-values have undergone no correction (raw data); on the right part, they have been subjected to the peak-based correction. Reproduced with permission from [386].

The divergent β value distributions preclude analysis of all probes together and indicate that these two types of probes should be considered separately in any analyses. However, several methods have been proposed for dealing with this probe design bias in order to enable analyses of the two probe designs together. The first method proposed was Peak Based Correction (PBC)[386]. PBC rescales the distribution of methylation values from type II probes on the basis of the distribution of methylation values from type I probes using the peaks of the bimodal distribution of methylation. This rescaling moves the methylation values corresponding to the unmethylated state closer to zero and the values corresponding to the methylated state more towards 1 [386].

PBC does not use β-values as methylation values but M-values. M-values are the log2 ratio of the intensities of methylated probe versus unmethylated probe, and are calculated using the equation:

$$M = \log_2 \left[(\max(M,0) + \alpha) / (\max(U,0) + \alpha)\right]$$

where M is the intensity of the methylated signal, U the intensity of the unmethylated signal at each probe and α an offset value (default α=1) to prevent small estimation errors translating into big changes [387]. An M value close to 0 shows that the methylated and unmethylated intensities are similar, therefore the CpG site has a methylation 50%. Positive M values show that more molecules are methylated than unmethylated (M>U), while negative values mean that the unmethylated intensity is stronger than the methylated intensity (U>M). PBC and other processing methods for the Infinium HumanMethylation 450 BeadChip make use of M-values instead of β-values because a) β-values are severely compressed at the extremes when compared with M-values, therefore the bimodal distribution is less obvious and b) β-values have significant heteroscedasticity at the extremes of the methylation distribution, whereas M-values are approximately homoscedastic making them more appropriate for the assumptions of the majority of statistical models used for microarray analysis [387].

However, PBC has been questioned for several reasons. The first is that it seeks to re-scale type II probe intensity/methylation distribution to match the distribution of type I probes. However, the different distributions might not be a performance issue but instead a consequence of the fact that the two probe designs measure different positions on the genome while making different assumptions. For example, type I probe design assumes that all CpGs on the probe body have the same methylation as the target CpG, whereas type II probe design makes no assumption about the methylation of the CpGs on the probe body (uses degenerate R bases corresponding to the C of each CpG). Also type II probe bodies can contain up to 3 CpG sites (the majority less than 3) whereas most type I probes contain 3 or more CpG sites. Lastly, due to their higher CpG content, type I probes are more likely to map to CpG dense regions such as CpG islands (57%) in contrast to type II probes only 21% of which are located in CpG islands. Given all these it is likely that the difference in methylation distributions between type I and type II probes is a true biological effect rather than a technical artefact [388] and PBC results in overcorrection or suppression of type II probe methylation values possibly masking biologically significant signals[389].

A second limitation of PBC is that it has been shown to break down when methylation does not exhibit a well-defined bimodal distribution, because PBC uses the peaks of the bimodal distribution to re-scale type II probe to type I probe methylation distribution [389]. This is especially true when measuring the methylation of tumour tissues since the aberrant methylation pattern observed in cancer will most likely violate both the expected bimodal distribution and the assumption of type I probes, namely that CpGs on the probe body will have the same methylation as the target CpG. If the latter assumption does not hold then the methylation of CpGs mapping to type I probes might be incorrectly assessed and normalization to type I probes will add further bias [390]. Lastly, as PBC tries to force the type II probe distribution on the type I probe distribution, it causes an “artificial expulsion” of data from the intermediate methylation region to the extremes of the distribution – both the methylated and the unmethylated extremes [389].

To satisfy the need for normalisation and to deal with the issue of differential probe design without the limitations of PBC, Subset Quantile Normalisation (SQN) and Subset-quantile Within Array Normalisation (SWAN) were proposed. Both methods are applied on signal intensities and then the normalised signal intensities are used

to calculate β-values. Also both methods are within sample/array normalisation methods, meaning that each sample is normalised independently. SQN uses type I probes as anchor probes and rank equivalence to normalise type II probes. First, type I and type II probes are ranked separately, and the non-anchor probes at the q^{th} percentile are made to have the same value as the anchor probes at that same percentile. The normalisation is performed separately for each probe category with respect to the "relation to CpGs" (S shore, S shelf, N shore, N shelf and distant) and the "relation to gene sequence" (body, 5'UTR, 3'UTR, 1^{st} exon, intergenic, multiple annotations, TSS200, and TSS1500) to account for the difference in CpG content and mapping of the two probe designs [391]. Despite being a within sample normalisation method, SQN also corrects for between sample variation (batch effects).

SWAN uses probes with a similar CpG content to define an average quantile distribution, as it assumes that probes with the same CpG content will produce similar intensity distributions reflecting their similar methylation due to their similar position in the genome. These probes can be a random sample of both type I and type II probes as long as the probe bodies contain the same number of CpG sites. Subsequent processing is performed independently on both methylated and unmethylated channels. For each random sample of probes with n CpG sites, type I and type II probes are separately sorted by increasing signal intensity and the mean intensity of the two probe types for each quantile is calculated. Then, the intensities of the remaining probes (type I and type II treated separately) are adjusted by interpolation of the distribution of the subset probes. This means that even though subset distributions are identical, the intensity distributions of type I and type II are not [388].

More recently, another method to normalise type I and type II probes has been developed that unlike PBC does not depend on a bimodal distribution of methylation and unlike SQN and SWAN makes no assumptions about methylation's dependence on probe CpG density and genome position. Beta Mixture Quantile Normalisation (BMIQ) adjusts the β-values of type II probes into a statistical distribution similar to that of type I probe β-values, by applying a beta mixture three-state methylation model. This model separates type I and type II probes independently into unmethylated (<20% methylation), hemi-methylated (20-80% methylation) and methylated (>80% methylation) probes. For type II probes within each group, the probability of belonging to that methylation group is converted into quantiles using the inverse of the cumulative beta distribution, with beta parameters estimated from the type I probes of the same methylation group [389]. Thus type II probes retain their proportions of un/hemi/methylated probes after normalization. In this way, BMIQ manages to respect the biological variability of type I and type II probes, it reduces the bias that comes from matching the density distributions of the two probe types at the unmethylated and methylated extremes, and it allows beta values of type II probes to retain their ranking [389].

2.2.3.1.3 Batch Effect Correction

Probably the last issue with Infinium HumanMethylation 450 BeadChip data that needs to be addressed before processing is the issue of batch effects. Batch effects describe any form of intra-experiment variation, usually time

and place-dependent, in experiments involving processing of several microarrays [392]. Leek et al., [393] demonstrate that batch effects in array data such as the 450K ILLUMINA can be substantial and these effects are not always removed by normalisation and can have severe consequences in the produced results. Uncorrected batch effects can lead to increased variability and decreased power to detect real biological signals, or if the sources of technical variation correlate with biological effects they can generate false positives. Thus it is important to look for and correct for batch effects in whole genome methylation statistical analyses.

If variation is caused by known simple surrogates, such as chip, time, and date of processing, then these variables can be entered in the statistical models as confounders [393]. However, when analyses do not involve statistical models, batch effects from known simple surrogates can be corrected for using ComBat, an empirical Bayes method. ComBat utilises Location and Scale (L/S) adjustments. L/S adjustments were popular for adjustment of batch effects even before ComBat and they assumed a model for the location (mean) and/or scale (variance) of data (each probe separately) within batches and then each batch is adjusted to meet this model specifications [394]. In other words, the mean and/or variance of each probe are standardised across batches in order to model out batch effects. In ComBat, model parameters that characterise the batch effects are estimated "by pooling information *across [probes]* in each batch to shrink the batch effect parameter estimates towards the overall mean of the batch effect estimates (across [probes]). These estimates are then used to adjust the data for batch effects"[394]. By pooling data across probes, ComBat does not require large sample sizes as it is robust to outliers [394].

If, however, the known surrogates are not enough to explain all technical variability and sources of batch effects are unknown, methods such as Surrogate Variable Analysis (SVA) can be used. SVA estimates the sources of batch variability directly from the data and generates surrogate variables that best capture batch effects using a Single Value Decomposition (SVD) algorithm. The SVA surrogate variables, which are linearly uncorrelated, can then be incorporated into statistical models as confounding variables [395]. A variation of SVA, Independent Surrogate Variable Analysis (ISVA) has been also proposed which models the surrogate variable as statistically independent and not just linearly uncorrelated, using Independent Component Analysis (ICA). ISVA has been shown to perform better, especially in larger datasets [396]. However, it is important to note that SVD and ISVA require many samples per batch (>25) to be implemented successfully [394].

The choice of pre-processing steps has been one of the most challenging parts of this thesis, as a variety of methods seems to perform equally well on this data and since the literature on which methods might be more appropriate is growing exponentially. As a result, a publication that compared pre-processing methods in order to derive the best performing combination of dye bias, normalisation and batch effect methods, was used to guide this analysis.

Marabita et al. [397] performed an evaluation of analysis pipelines, specifically using the Illumina HumanMethylation450 BeadChip platform. The metrics assessed were a) reduction of technical variability during normalisation using replicates, b) removal of probe design bias using degree of matching of methylation

distributions between probe types as well as reduction in methylation variability between adjacent probes with differential design, c) efficacy on removing batch effects using Principal Component Analysis (PCA) and clustering techniques, and d) ability to detect differentially methylated sites. Six different combinations of pre-processing steps were investigated. Overall, after probe and sample filtering, the most effective combination of normalization methods was shown to be:

- Dye/Colour bias adjustment
- QN on signal intensities
- BMIQ normalization
- ComBat.

However, it was also shown that the same pipeline without QN performed equally well [397]. Given these results, the same combination of pre-processing methods was decided to be used on the data, without QN on signal intensities.

Dye bias correction was performed by the bioinformatician of the group, Gianluca Campanella, as previously mentioned. BMIQ was performed on each sample in the dataset using the *BMIQ* command in the wateRmelon package. Dye bias correction and BMIQ are within-sample normalization methods therefore they were applied to the whole dataset. ComBat was performed next, using the *combat* command in package CH_3, an R package created by the bioinformatician of the group. The input was a matrix of dye bias adjusted and BMIQ normalized β-values, but the command itself transforms β-values into M-values in order to correct for batch effects. The largest variation in this dataset was observed across chips, therefore, for ComBat, each chip was considered as a different batch. Since ComBat is a between array normalization method, it was performed first on the entire dataset and then independently on a smaller subset of the dataset including only female subjects, the "females only" dataset. This was because the entire dataset was used to investigate exposures common to males and females whereas the smaller subset was used for the investigation of those exposures only relevant to women.

2.2.3.1.4 Probe Exclusions

Lastly, for whole genome methylation and clustering analyses, several compromising probes had to be removed because their design was imperfect and could have confounded methylation readings and hence analyses results. To begin with, probes interrogating 50 bp regions containing Single Nucleotide Polymorphisms (SNPs) were excluded from the analysis because SNPs in the probe body could affect hybridization and hence methylation measurements. In addition, expanded biological annotation of the array revealed that 4.3% of the probes had a CpG target site which was a documented SNP and 32.5% of these were not labelled as containing a SNP in the array annotation files[398]. These probes allow sample genotype to confound the methylation measurement, hence they were also excluded. Furthermore, 8.6% of probes mapped to multiple locations of the genome – between two and five – as assessed *in silico* [398]. The majority of these were intended to target autosomal loci but mapped to sex chromosomes as well, whereas a small number were intended to bind to sex loci and also mapped to autosomal chromosomes. Still a large percentage of these cross-hybridizing probes contained a sequence of base pairs that in its entirety matched and could bind to repetitive regions of the genome [398]. Probes containing SNPs in their body or target site as well as cross-hybridizing probes were identified and excluded from the analyses based on the additional annotation files made public on NCBI Gene Expression Omnibus (GEO) by Price et al. [398], under the title "Illumina HumanMethylation450 BeadChip [UBC enhanced annotation v1.0]" Platfrom GPL16304.

The SNP containing probes were excluded mainly because of the observed increased variation in β-value between individuals for probes with a SNP either in their probe body or on their target CpG compared to probes without SNPs[398,399]. Given the small per probe methylation differences observed between individuals the increased variability of the SNP containing probes could identify these probes as most influential and they would heavily influence clustering and other "probe-grouping" analyses. The same possibility stands for cross-hybridizing probes since even if only a few probes displayed erroneous methylation readings because of cross-hybridization, this would be enough to significantly bias clustering and other analyses involving grouping of probes. Furthermore, probes matching to the sex chromosomes were excluded from whole genome and clustering analyses since they would inevitably be differentially methylated between males and females and this could lead to gender bias.

Lastly, the importance of taking the type of different WBCs into account when analysing DNA methylation in whole blood has been emphasized in a number of publications [400,401]. However, WBC differentials were not available for the samples analysed here. In order to address this issue of tissue heterogeneity, 450K BeadChip methylation data obtained from purified CD4 T-cells, CD8 T-cells, CD19 B-cells, monocytes, natural killer (NK) cells, neutrophils, eosinophils and whole peripheral blood mononuclear cells (PBMCs) in an independent investigation[401] were used. The data was analysed by a colleague, Dr. James Flanagan, to identify probes whose methylation differed significantly between each individual cell type and PBMC (linear regression using β values, $p<1e-07$ and delta-β >0.05). This identified n=11,196 unique probes, which were removed from the whole genome methylation and clustering analyses, assuming that blood composition only marginally affected methylation patterns at other sites (remaining probes).

However, given the large number of probes that were removed based on the additional annotation, XY position, and differential WBC methylation, in order to avoid blindly excluding a large portion of this dataset, all these probes were retained in the locus by locus analyses and the significantly differentially methylated probes between groups were checked to ensure that the variation was a true biological difference and not an artefact of imperfect probe design, gender bias, or tissue heterogeneity. In addition, as a further safeguard, all locus by locus analyses were also performed without these probes to ensure that the results did not differ.

2.2.3.2 Whole Genome Methylation analysis

In order to replicate the results observed with LUMA, the median beta value across all probes was calculated per individual to derive a measure of whole genome methylation per subject. Similarly, in order to determine which probes contribute most to differential methylation between exposure groups, the median beta values across probes in different Functional Genomic Distribution (FGD) and CpG island (CGI) and neighbourhood contexts, were calculated per individual. More specifically the following FGD and CGI context categories were used to calculate probe type specific median methylation values per subject: Promoter (including probes in regions TSS200, TSS1500, 5'UTR, and 1st Exon), gene body, 3'UTR, intergenic, CPG Island, CGI shore, CGI shelf and other ("open sea") as described by Sandoval et al. [402]. Lastly, probes that accidentally matched to repetitive elements (and were excluded from the above-mentioned categories, as previously described) were separately grouped together per individual to get a very rough and debatable estimate of repetitive element methylation per subject. Linear regression was used to examine whether whole genome methylation and methylation at each of these probe category specific groups was significantly different between subjects in different exposure categories. The variables examined for each exposure of interest are presented in Table 7 and the confounding variables entered into each regression model are the same as those used for the locus by locus analyses (Table 8). For each variable examined, tests were carried out twice – once on all subjects ($n_{\text{both sexes}}$=667 and $n_{\text{females only}}$=485) and once on only those subjects that remained healthy for at least five years following blood collection ($n_{\text{both sexes}}$=504 and $n_{\text{females only}}$=357).

Table 7 – Variables examined for each exposure of interest

Exposure of Interest	Subjects	Variables	Variable Categories
Adult SES	ALL (n=670)	Education 4 categories	University
			Secondary School
			Technical / Professional School
			Primary School or no education
		Education 2 categories	College
			Lower than College
		House Density	<1 room/person
			≥ rooms / person
		Household's highest occupational position	High (white collar occupation)
			Middle (retailers, craftsmen)
			Low (blue collar workers and farmers)
Childhood SES	ALL (n=670)	Father's occupational position	Manual
			Non-manual
		Birth during WWII	No (<1939 and >1944)
			Yes (1939 – 1944)
		When household first had fridge	<1954
			1955-1964
			>1964
Nutrition	ALL (n=670)	Italian Mediterranean Diet Index	Score categories: 0-3 = Low adherence 4-6 = Moderate adherence 7-10 = High adherence
		Total Caloric Intake	Intake in quartiles: Q1 = ≤1713.7/day Q2 = >1713.7 - ≤2104.2/day

			Q3 = >2104.2 - ≤2539.7/day Q4 = >2539.7/day
		Alcohol intake	Intake in quartiles: Q1 = ≤0.30g/day Q2 = >0.30 - ≤5.65g/day Q3 = >5.65 - ≤22.7g/day Q4 = >22.7g/day
		Folate intake	Intake in quartiles: Q1 = ≤216.3 µg/day Q2 = >216.3 - ≤268.5 µg/day Q3 = >268.5 - ≤325.4 µg/day Q4 = >325.4 µg/day
		Methyl Intake Score	Score categories: 0: no 1-2: low 3-4: moderate 5-6: high methyl intake.
Overweight / Obesity	ALL (n=670)	BMI	Categories: <25 kg/m^2 25-30 kg/m^2 ≥30 kg/m^2
Height	ALL (n=670)	Height	Height Quartiles: Q1 = ≤156 cm Q2 = >151 - ≤161 cm Q3 = >161 - ≤167.5 cm Q4 = >167.5 cm
Cumulative oestrogen exposure	Women only (n=488)	Age at menarche	Categories: ≤11 years 12-14 years ≥15 years
	Women only (n=488)	Menopausal Status	Categories: Pre-menopausal Post-menopausal
	Post-menopausal women (n=307)	Age at menopause	Categories: ≤50 years >50 years
	Women only (n=488)	Estimated Lifetime oEstrogen Exposure (ELEE)	ELEE Quartiles: Q1 = ≤32 years Q2 = >32 - ≤36 years Q3 = >36 - ≤39 years Q4 = >39 years

2.2.3.3 Genome wide locus by locus Methylation analysis

Locus by locus analysis of methylation was performed using a linear regression model. In order to satisfy the normality assumptions of a linear regression model, beta methylation values were converted to M-values as described above and entered into the model as a continuous outcome. Each exposure of interest was investigated using the variables in Table 8 as categorical and/or continuous. All models were adjusted for age at blood collection, case-control status and position on the chip which in this dataset was shown to confer a bias on the methylation measurement. Additional confounders such as gender, smoking status, BMI, total caloric intake and folate and alcohol intakes were entered into the models where relevant as shown in Table 8. All confounders were entered into the models as continuous variables with the exception of chip position and smoking status. Analyses were performed first on all relevant subjects ($n_{\text{both sexes}}$=667 and $n_{\text{females only}}$=485) and then repeated only on the subjects that remained healthy for at least five years following recruitment ($n_{\text{both sexes}}$=504 and $n_{\text{females only}}$=357). For the reproductive variables, all of which have been consistently shown to be strongly associated with breast cancer, analyses were also performed on only the female subjects who remained free of cancer for the entire length of follow-up which was several years ($n_{\text{female controls only}}$=243).

Q-values, measuring the maximum False Discovery Rate (FDR) from the Benjamini and Hochberg method were derived for each probe analysis, and overall Type I error was controlled for by conservatively applying Bonferroni multiple testing correction. The per-test significance Bonferroni corrected threshold was set to 1.00×10^{-7} with a corresponding FDR cut off of 0.05. For each probe investigated, given the available sample size, a significance level of 0.05 and a power of 80%, only methylation differences >28% between exposure groups can be considered with confidence.

Even though a candidate gene approach would minimize the problem of multiple testing, given the limited available knowledge about the effects of such exposures on the epigenome and the lack of knowledge regarding locations of exposure relevant variation[403], it was decided to maintain a genome wide rather than a gene-specific approach in this exploratory investigation.

Table 8 – Variables examined and confounders adjusted for in linear regression models with methylation as continuous outcome

Exposure of Interest	**Variables**	**Variable Categories**	**Confounders adjusted for in addition to age at blood collection, case-control status, and position on the chip**
Adult SES	Education 4 categories	University	Gender, smoking status, BMI, and daily alcohol and folate intakes
		Secondary School	
		Technical / Professional School	
		Primary School or no education	
	Education 2 categories	College	Gender, smoking status, BMI, and daily alcohol and folate intakes
		Lower than College	
	House Density	<1 room/person	
		≥ rooms / person	
	Household's highest	High (white collar occupation)	Gender, smoking Status, BMI,

	occupational position	Middle (retailers, craftsmen)	and daily alcohol and folate intakes
		Low (blue collar workers and farmers)	
Childhood SES	Father's occupational position	Manual	Gender
		Non-manual	
	Birth during WWII	No (<1939 and >1944)	Gender
		Yes (1939 – 1944)	
	When household first had fridge	<1954	Gender
		1955-1964	
		>1964	
Nutrition	Italian Mediterranean Diet Index	Score - Continuous	Gender, education, total caloric energy intake, BMI, and physical activity
		Score categories: 0-3 = Low adherence 4-6 = Moderate adherence 7-10 = High adherence	
	Total Caloric Intake	Intake in quartiles: Q1 = /day Q2 = /day Q3 = /day Q4 = /day	Gender, education, BMI, physical activity, folate and alcohol intakes
	Alcohol intake	Intake in quartiles: Q1 = ≤0.30 g/day Q2 = >0.30 - ≤5.65g/day Q3 = >5.65 - ≤22.7g/day Q4 = >22.7g/day	Gender, education, total caloric energy intake, folate intake, BMI, and physical activity
	Folate intake	Intake in quartiles: Q1 = ≤216.3 µg/day Q2 = >216.3 - ≤268.5 µg/day Q3 = >268.5 - ≤325.4 µg/day Q4 = >325.4 µg/day	Gender, education, total caloric energy intake, alcohol intake, BMI, and physical activity
	Methyl Intake Score	Score categories: 0: no 1-2: low 3-4: moderate 5-6: high methyl intake	Gender, education, total caloric energy intake, BMI, and physical activity
Overweight / Obesity	BMI	Continuous	Gender, total caloric energy intake, percent energy intake from fat and sex adjusted total physical activity
		Categories: <25 kg/m^2 25-30 kg/m^2 ≥30 kg/m^2	
Height	Height	Height Quartiles: Q1 = ≤156 cm Q2 = >151 - ≤161 cm Q3 = >161 - ≤167.5 cm Q4 = >167.5 cm	Gender, total caloric energy intake
Cumulative oestrogen exposure	Age at menarche	Continuous	BMI
		Categories: ≤11 years 12-14 years ≥15 years	
	Menopausal Status	Categories: Pre-menopausal Post-menopausal	
	Age at menopause	Continuous	
		Categories: ≤50 years >50 years	
	Estimated Lifetime oEestrogen Exposure (ELEE)	Continuous	
		ELEE Quartiles: Q1 = ≤32 years	

		Q2 = >32 - ≤36 years Q3 = >36 - ≤39 years Q4 = >39 years	

2.2.3.4 Clustering

Clustering was also performed to further investigate whether exposure variables have a widespread influence on methylation patterns and not just single loci. The goal of clustering was to assign subjects to classes/groups on the basis of methylation measurements so that subjects in a group/class have a similar methylation pattern with respect to the probes used for clustering. Then, whether the clusters formed were related to the exposure of interest was examined.

For each dataset – all subjects and "females only" – the 1000 most variable probes were used to cluster subjects based on their methylation profile. Variability was assessed by calculating the standard deviation for each probe. For clustering, the 1000 most variable probes were identified after exclusion of SNP containing probes, probes with compromised design, and XY probes, since these probes would inevitably be amongst the most variable and would bias clustering.

Initially, for each examined exposure unsupervised hierarchical clustering (HC) of subjects was performed based on methylation of the most variable probes using Euclidean distance and average linkage for the separation of clusters. Hierarchical clustering "builds clusters by repeatedly joining and merging those objects separated by the shortest distance. After the two closest objects are merged, the distance matrix is updated and the process is repeated to merge all objects" [404]. HC results are displayed in a dendrogram where distance on the y-axis represents distance between clusters and observations. In contrast to partitioning cluster methods, in HC there is no need to define a-priori the number of clusters to be produced. However, in order to carry out clustering, distance and linkage need to be defined. Distance is a metric of the degree of similarity or dissimilarity between subjects within a cluster. Linkage defines the between-cluster dissimilarity and it is used when deciding the distance between groups of observations i.e. from which point in a cluster to measure the inter-cluster difference [404]. Euclidean distance is the most commonly used distance and creates sphere shaped clusters. Average linkage uses the mean distance between subjects in different clusters. Hierarchical clustering of subjects based on methylation of the most variable probes was visually displayed in heatmaps created using the package *ggplots2* in R. Using the heatmaps, whether different exposure groups differed in their methylation profile was visually examined.

Despite HC based on Euclidean distance with average linkage being the most popular clustering method, the results of a recent comparison of HC and partitioning clustering methods by Clifford et al. on simulated data led us to also investigate other clustering methods [404]. The analysis determined that in large datasets where a large proportion of probes used are not differentially methylated between groups – conditions that are true in this case – the best metric to separate clusters seems to be Manhattan distance with a k-means method [404]. K-means is not a hierarchical but a partitioning clustering method that re-allocates observations to clusters so that each observation is added to the cluster with the nearest mean. Manhattan distance creates clusters with diamond shapes and has been shown to be more robust to outliers [404].

In order to investigate this method of clustering, Consensus Clustering (CC) was performed using the package *concensusClusterPlus* in R. Consensus clustering is also an unsupervised class discovery tool, and its aim is to detect classes of items similar with respect to particular intrinsic features [405], in this case DNA methylation. CC runs a clustering algorithm multiple times and provides evidence which allows the user to determine the most robust number of clusters (k) and to assess the stability of the clusters formed. The multiple runs can also have different random starting points since the basic assumption is that different samples drown from the same sub-population, should induce a similar cluster arrangement[405].

The algorithm used by CC can be specified by the user depending on the needs of each analysis. In this investigation, based on the results of Clifford et al.,[404] a k-means algorithm was used with Manhattan as the distance metric. The maximum number of clusters was specified as 10 and the clustering algorithm was re-run on 100 different, randomly selected subsets, each consisting of 80% of the samples in a dataset. The advantage of CC was that in contrast to other partitioning clustering methods, the number of clusters did not need to be pre-specified. Instead, results of clustering of multiple sub-samples, for each different number of resulting clusters (k), were graphically visualized, enabling decisions to be made concerning the most appropriate cluster number and cluster membership [405]. As a general rule, more confidence was placed on cluster structures that were robust to sampling variability [405].

In order to determine robustness of clustering results, the proportion of clustering runs in which two observations were grouped together, termed as *pairwise consensus values*, were stored in a *consensus matrix* and imaged as heatmaps. One heat map was produced for each resulting number of clusters, for example 1 heatmap for k=2, another for k=3, another for k=4 and so forth until k=10. These heat maps were used to determine the stability of clustering for each k, by identifying the cleanest cluster partition where there is almost complete concordance between sub-sample clustering runs– meaning that subjects are either always clustered together (dark blue colour) or are always not clustered together (white colour). In addition, empirical cumulative distribution function (CDF) plots for each k, estimated from histograms of 100 bins, were used to determine at which k the distribution approaches maximum, which indicates maximum stability [405]. The membership of each subject in the final *consensus clusters* for each k was determined using agglomerative hierarchical consensus clustering using a distance of 1-*pairwise consensus values*[405]. As an example, for a given k, if the proportion of times two subjects were grouped in the same cluster was high, then they would have a high *pairwise consensus value*, so that in clustering using a distance of 1-*pairwise consensus values* they would be close together in the same cluster.

Once the most robust number of clusters based on the 1000 most variable probes was determined, the association between cluster membership and the exposure variables of interest was investigated using chi-square tests.

2.2.3.5 Principal Component Analysis

In order to characterize the main features distinguishing participants included in this study and to determine which features explain most of the methylation variance in our data, principal component analysis (PCA) was performed on the samples. The input for the PCA was a matrix with the samples/subjects as columns and the methylation across all 444,278 probes as rows. Principal Component Analysis was performed once on the entire dataset including all subjects (n=667) and the resulting principal components (PCs) were regressed against variables relevant to both genders, namely education (in 4 and 2 categories), household density, highest combined household occupational position, father's occupation, date at which house first had a fridge, birth during the war, IMI score, caloric intake, alcohol intake, folate intake, dietary methyl intake, height, and BMI. The association of these components with chip, age, gender and case-control status was also investigated. PCA was then also run on the "females-only" dataset (n=485), and the resulting PCs were regressed against age at menarche, menopausal status, age at menopause, ELEE, chip, age, and case-control status. All categorical exposure variables were treated as continuous in the regression models, attempting to examine the effect of change in exposure category on the PC.

The first PC obtained from PCA on all subjects explained 99.06% of the variance in methylation, while the first component from PCA on the female subjects explained 99.16% of the variance in methylation. This suggests that heterogeneity in the study population can be accurately summarized using a single axis. However, for the sake of completeness the two subsequent principal components were alse described and regressed against the exposures of interest.

2.2.3.6 Single Enrichment Analysis

For those exposures suspected to have a widespread influence on methylation, Single Enrichment Analysis (SEA) was performed on probes which were found to be differentially expressed between exposure categories, at the $1.0x10^{-5}$ significance cut off point. Single Enrichment Analysis was performed to identify whether there was an over-representation of genes with specific functional annotations in the selected gene lists. SEA was performed using the DAVID software as described in the associated protocols [406,407]. According to the recommendations in the protocols, only exposures with at least 100 differentially methylated probes at the pre-set threshold were examined. However, given the relaxed threshold used ($1.0x10^{-5}$ vs. $1.0x10^{-7}$) SEA was performed as an exploratory tool rather than a discovery tool.

Gene functional classification, as well as functional annotation clustering, were performed using the highest stringency for the fuzzy algorithms and only clusters with an enrichment score ≥1.3 were deemed as significant. For functional annotation, biological terms were considered only if their P-value was below the Benjamini threshold of 0.05.

3 Results

3.1 Global Methylation (LUMA)

Demographic, anthropometric, lifestyle, and reproductive characteristics for subjects with LUMA global methylation measurements are presented in Table 9. The median genome-wide methylation in these subjects was 71.7% and the standard deviation was 5.7%. Of all the anthropometric measures and lifestyle variables examined, only age at menarche was found to significantly differ across quartiles of percent global methylation. Higher genome wide methylation was associated with a younger age at menarche (Kruskal-Wallis P-value=0.002), and this association was significant even after Bonferroni correction for multiple testing (Table 9). Age at blood collection, height, weight, BMI, physical activity, smoking status, daily alcohol, folate consumption, Mediterranean diet adherence score, age at FFTP, menopausal status, age at menopause, ELEE, parity, breastfeeding, oral contraceptive (OC) use, hormone replacement therapy (HRT) use and highest level of education achieved did not significantly differ between subjects in methylation quartiles (Table 10).

Table 9 – Demographic, anthropometric, lifestyle, and reproductive characteristics for subjects with LUMA measurements

Covariate			**No. Subjects with available information**
Age	Range	33.4 – 75.6	
	Median	52.7	
	Mean (sd)	52.9 (9.4)	376
Birth during the WWII:			
- **No**	n (%)	324 (86.2)	
- **Yes**	n (%)	52 (13.8)	376
Highest Educational Level			
- **None**	n (%)	29 (8.0)	
- **Primary**	n (%)	157 (43.1)	
- **Technical/Professional school**	n (%)	62 (17.0)	
- **Secondary**	n (%)	67 (18.4)	
- **University**	n (%)	49 (13.5)	364
Height	Range	136.8 – 185	
	Median	160	
	Mean (sd)	160.1 (6.7)	376
Weight	Range	39.6 – 110.2	
	Median	64.5	
	Mean (sd)	66.2 (11.2)	376
BMI[±]			
- **<25 kg/m^2**	n (%)	182 (48.4)	
- **25-30 kg/m^2**	n (%)	141 (37.5)	
- **≥30 kg/m^2**	n (%)	53 (14.1)	376
Waist-to-hip ratio	Range	0.65 – 0.99	
	Median	0.79	
	Mean (sd)	0.79 (0.06)	376
Waist Circumference	Range	61 – 120.5	
	Median	80	

	Mean (sd)	81.2 (10.8)	376
Physical Activity			
- **Inactive**	n (%)	34 (9.0)	
- **Moderately Inactive**	n (%)	85 (22.6)	
- **Moderately Active**	n (%)	210 (55.9)	
- **Active**	n (%)	45 (12.0)	
- **Missing**	n (%)	2 (0.5)	376
Smoking Status			
- **Current Smoker**	n (%)	79 (21.1)	
- **Former Smoker**	n (%)	65 (17.4)	
- **Never**	n (%)	230 (61.5)	376
Daily energy intake (calories/day)	Range	950.4 – 4112.3	
	Median	1876.7	
	Mean (sd)	1947.1 (537.0)	376
Daily alcohol consumption (g/day)	Range	0 – 51.2	
	Median	3.5	
	Mean (sd)	6.5 (8.4)	362
Daily folate consumption (µg/day)	Range	90.4 – 1113.0	
	Median	268.5	
	Mean (sd)	291.5 (107.9)	376
Mediterranean Adherence Score	Range	0 – 9	
	Median	4	
	Mean (sd)	4.2 (1.8)	376
Methyl Intake Score	Range	0 – 6	
	Median	3	
	Mean (sd)	3.0 (1.8)	376
Age at Menarche			
- **≤ 11 yrs**	n (%)	72 (19.4)	
- **12 – 14 yrs**	n (%)	242 (65.2)	
- **≥ 15 yrs**	n (%)	57 (15.4)	371
Age at FFTP[±]			
- **<25 yrs**	n (%)	147 (45.6)	
- **25 – 30 yrs**	n (%)	131 (40.7)	
- **≥30 yrs**	n (%)	44 (13.7)	322
Parous			
- **No**	n (%)	45 (12.0)	
- **Yes**	n (%)	330 (88.0)	375
Breastfeeding			
- **No**	n(%)	109 (30.0)	
- **Yes**	n(%)	255 (70.0)	364
Menopausal Status			
- **Premenopausal**	n (%)	175 (46.5)	
- **Postmenopausal**	n (%)	201 (53.5)	376
Age at menopause			
- **≤50 years**	n (%)	123 (63.7)	
- **>50 years**	n (%)	70 (36.3)	193
ELEE	Range	16 – 49	
	Median	34	
	Mean (sd)	33.8 (5.1)	365
HRT[±] **use**			
- **Ever**	n (%)	29 (7.8)	
- **Never**	n (%)	343 (92.2)	372
OC[±] **use**			
- **Ever**	n (%)	176 (46.9)	
- **Never**	n (%)	199 (53.1)	375

± BMI: Body Mass Index, FFTP: First Full Term Pregnancy, HRT: Hormone Replacement Therapy, OC: Oral Contraceptive

Table 10 – Anthropometric and lifestyle variables in healthy subjects with respect to LUMA methylation quartiles

Variable		Methylation Quartile: Cut-offs				
	Units	1: 23.0-68.4 % (n=103)	2: 68.5-71.7 % (n=91)	3: 71.8-74.0 % (n=93)	4: 74.1-80.0 % (n=89)	P-value[a]
Age at blood collection	Mean ± SD	52.9 ± 8.5	52.2 ± 8.5	53.1 ± 10.0	53.6 ± 10.7	0.863
Birth during WWII[±]	N, (%)					
- No		85 (82.5)	76 (83.5)	84 (90.3)	79 (88.8)	
- Yes		18 (17.5)	15 (16.5)	9 (9.7)	10 (11.2)	0.316
Height	Mean ± SD	159.0 ± 6.1	160.7 ± 6.4	160.8 ± 6.7	160.0 ± 7.5	0.424
Weight	Mean ± SD	66.0 ± 12.3	66.5 ± 10.5	64.8 ± 11.3	67.7 ± 10.4	0.316
BMI [±]	Mean ± SD	26.2 ± 5.1	25.8 ± 4.1	25.1 ± 4.4	26.5 ± 4.4	0.112
Waist-to-hip ratio	Mean ± SD	0.80 ± 0.06	0.78 ± 0.06	0.79 ± 0.06	0.80 ± 0.06	0.514
Waist circumference	Mean ± SD	81.7 ± 11.5	80.3 ± 10.6	79.9 ± 10.6	82.8 ± 10.2	0.439
Physical Activity	N, (%)					
- Inactive		4 (3.9)	9 (9.9)	11 (11.8)	10 (11.2)	
- Moderately Inactive		29 (28.2)	22 (24.2)	16 (17.2)	18 (20.3)	
- Moderately Active		60 (58.3)	45 (49.4)	54 (58.1)	51 (57.3)	
- Active		9 (8.7)	15 (16.5)	11 (11.8)	10 (11.2)	
- Missing		1 (0.9)	0 (0.0)	1 (1.1)	0 (0.0)	0.404
Smoking Status	N, (%)					
- Never		58 (56.3)	59 (64.8)	49 (52.7)	47 (52.9)	
- Former		23 (22.3)	12 (13.2)	20 (21.5)	22 (24.7)	
- Smoker		22 (21.4)	20 (22.0)	23 (24.7)	19 (21.3)	
- Unknown		0 (0.0)	0 (0.0)	1 (1.1)	1 (1.1)	0.596
Daily energy intake (calories/day)	Mean ± SD	1985.2 (588.6)	1990.6 (535.7)	1929.6 (534.0)	1876.7 (476.2)	0.545
Alcohol consumption – lifetime average (g/day)	Mean ± SD	7.0 ± 8.6	6.9 ± 9.7	5.4 ±6.4	6.7 ±8.6	0.822
Dietary folate intake (g/day)	Mean ± SD	294.6 ± 121.9	276.9 ± 96.7	293.1 ± 104.2	301.0 ± 105.3	0.386
Mediterranean Adherence Score	Mean ± SD	4.07 (1.8)	4.35 (1.7)	4.30 (2.0)	3.96 (1.9)	0.406
Methyl Intake Score	Mean ± SD	3.0 (1.9)	2.8 (1.9)	3.0 (1.8)	3.2 (1.7)	0.513
Age at menarche	Mean ± SD	13.2 ± 1.7	13.3 ± 1.7	12.5 ± 1.4	12.8 ± 1.7	**0.002***
Age at FFTP[±]	Mean ± SD	25.3 ± 3.5	26.2 ± 3.9	24.8 ± 4.0	25.0 ± 3.7	0.132
Menopausal Status						
- Pre	N, (%)	42 (40.8)	47 (51.6)	46 (49.5)	40 (44.9)	
- Post		60 (58.3)	43 (47.3)	45 (48.4)	49 (55.1)	
- Surgical Post		1 (0.9)	1 (1.1)	2 (2.1)	0 (0.0)	0.545
Age at Menopause	Mean ± SD	48.8 (4.6)	49.2 (4.3)	48.8 (5.1)	47.6 (5.7)	0.672
ELEE	Mean ± SD	33.9 (4.6)	33.8 (4.6)	34.2 (5.3)	33 (5.9)	0.760
Parous	N, (%)					
- No		12 (11.7)	7 (7.7)	13 (14.0)	19 (21.3)	
- Yes		90 (87.4)	84 (92.3)	79 (84.9)	70 (78.7)	
- Unknown		1 (0.9)	0 (0.0)	1 (1.1)	0 (0.0)	0.075
Breastfeeding	N, (%)					
- No		30 (29.1)	19 (20.9)	26 (28.0)	34 (38.2)	
- Yes		70 (68.0)	70 (76.9)	62 (66.7)	53 (59.6)	
- Unknown		3 (2.9)	2 (2.2)	5 (5.3)	2 (2.2)	0.086
OC[±] use	N, (%)					
- No		50 (48.5)	53 (58.2)	47 (50.5)	49 (55.1)	
- Yes		53 (51.5)	38 (41.8)	46 (49.5)	39 (43.8)	
- Unknown		0 (0.0)	0 (0.0)	0 (0.0)	1 (1.1)	0.512
HRT[±] use	N, (%)					
- No		93 (90.3)	84 (92.3)	81 (87.1)	85 (95.5)	
- Yes		3 (2.9)	6 (6.6)	10 (10.7)	4 (4.5)	
- Unknown		7 (6.8)	1 (1.1)	2 (2.2)	0 (0.0)	0.399
Highest Education	N, (%)					
- None		10 (9.6)	6 (6.6)	7 (7.4)	6 (6.8)	
- Primary		45 (43.7)	41 (45.1)	36 (38.7)	35 (39.3)	
- Technical/						

	Professional		15 (14.6)	9 (9.9)	18 (19.4)	20 (22.5)	
-	**Secondary**		12 (11.7)	21 (23.1)	17 (18.3)	17 (19.1)	
-	**University**		12 (11.7)	13 (14.3)	14 (15.1)	10 (11.2)	
-	**Unspecified**		9 (8.7)	1 (1.0)	1 (1.1)	1 (1.1)	0.264

[a] For continuous variables, P-value was derived from Kruskal-Wallis test. For categorical variables, P-value was derived from a chi square test, with the exclusion of "Unknown" categories due to their small cell counts. Both reflect the association between quartiles of methylation and the investigated variables.

*Significant at the Bonferroni-corrected significance cut off (P=0.003) for multiple comparisons

[±] WWII: World War II, BMI: Body Mass Index, FFTP: First Full Term Pregnancy, HRT: Hormone Replacement Therapy, OC: Oral Contraceptive

3.1.1.1 Socioeconomic Status

Highest level of education attained was not associated with global DNA methylation, despite adjustment for other relevant confounders such as centre, plate, case/control status, age at blood collection, smoking status, BMI, and folate and alcohol daily intakes (Table 11).

Table 11 – Association of highest level of education attained and birth during WWII with above/below median global methylation in a logistic regression model

	N (%)	Mean Methylation (SD)	Odds Ratio (OR)[¤]	95% Confidence Interval
Highest Level of Education Attained				
None	29 (8.0)	69.9(4.9)	1.00	
Primary School	157 (43.1)	70.1 (6.7)	1.04	0.45 – 2.38
Technical / Professional school	62 (17.0)	71.9 (4.4)	0.58	0.22 – 1.50
Secondary school	67 (18.4)	71.7 (3.7)	0.80	0.32 – 2.03
University	49 (13.5)	70.3 (6.3)	0.92	0.34 – 2.47
Birth during World War II				
No	324 (86.2)	70.7 (5.8)	1.00	
Yes	52 (13.8)	70.0 (5.2)	1.81	0.99 – 3.94

[¤] OR in a logistic regression model with above and below median methylation as the outcome, adjusting for centre, plate, case/control status, age at blood collection, smoking status, BMI, and folate and alcohol intakes.

Dividing subjects in two educational groups, those with and those without university education did not evidence significant results ($OR_{\text{university vs. no university education}}$: 1.29, 95%CI: 0.69–2.40). Similarly comparing those with any education against those with no education did not produce any association with global methylation ($OR_{\text{any education vs. no education}}$: 0.96, 95%CI: 0.43–2.17).

Despite the non-significant results on adult SES, birth during the war was borderline significantly associated with global hypomethylation (OR: 1.81 95%CI: 0.99 – 3.94, P=0.059) when adjusting for centre, plate, case/control status, age at blood collection, smoking status as well as BMI. It is important to note here that only a small number of subjects were born during WWII and the power of the analysis might not have been sufficient to produce a significant association.

3.1.1.2 Nutrition

Examination of the distribution of each food group (Table 12) and food nutrient (Table 13) between hypo- and hyper- methylated subjects indicated that intakes of single food groups or nutrients did not significantly differ between methylation groups, before adjusting for confounders.

Table 12 – Median and inter quartile range of daily intakes of specific food groups by percent global methylation group. P-values for Mann-Whitney Test.

Food Group (g/day)	Percent Global Methylation		Mann-Whitney P-value
	Below Median (N=194)	Above Median (N=182)	
	Intake Median (IQR)	Intake Median (IQR)	
Potatoes	55.72 (72.45)	55.84 (67.46)	0.843
Vegetables	161.32 (116.70)	166.32 (145.53)	0.476
Legumes	10.80 (16.08)	10.00 (19.33)	0.893
Fruit	275.66 (211.61)	252.85 (250.83)	0.628
Dairy	286.26 (217.81)	301.21 (295.30)	0.867
Cereal	172.84 (113.99)	163.80 (128.47)	0.442
Meat & Meat Products	90.07 (61.13)	89.25 (54.69)	0.749
Fish	22.46 (31.21)	22.05 (32.11)	0.160
Sugar & Sweets	32.86 (36.64)	30.35 (30.68)	0.189
Olive Oil	9.05 (23.02)	3.86 (20.44)	0.115

Table 13 – Median and inter quartile range of daily intakes of specific nutrients and energy by percent global methylation group. P-values for Mann-Whitney Test.

Food Group (g/day)	Percent Global Methylation Below Median (N=194)	Above Median (N=182)	Mann-Whitney P-value
	Intake Median (IQR)	Intake Median (IQR)	
Proteins (g)	82.98 (28.19)	81.41 (29.12)	0.336
Carbohydrates (g)	210.76 (80.15)	203.36 (75.78)	0.412
Fat (g)	73.60 (31.54)	71.22 (31.57)	0.312
Alcohol (g)	3.19 (11.77)	1.99 (10.59)	0.419
Cholesterol (g)	292.75 (164.60)	273.35 (172.45)	0.180
Fibre (g)	21.11 (9.22)	20.83 (8.88)	0.952
Vitamin C (mg)	124.37 (84.29)	123.62 (84.53)	0.823
Beta-carotene (µg)	2506.05 (1772.33)	2495.84 (2060.64)	0.978
Vitamin E (mg)	10.67 (5.91)	10.90 (6.97)	0.305
Potassium (mg)	3348.27 (1172.05)	3416.46 (1139.36)	0.372
Calcium (mg)	957.53 (481.17)	953.45 (440.15)	0.563
Phosphorus (mg)	1353.79 (552.42)	1389.17 (505.49)	0.916
Iron (mg)	11.57 (4.50)	12.09 (4.49)	0.851
Magnesium (mg)	291.90 (102.04)	301.24 (94.87)	0.468
Sugar (g)	98.10 (51.19)	94.97 (51.42)	0.414
Retinol (µg)	486.72 (559.28)	478.78 (562.12)	0.841
Riboflavin (mg)	1.55 (0.60)	1.58 (0.79)	0.341
Thiamin (mg)	1.05 (0.48)	1.06 (0.54)	0.456
Cobalamin (µg)	5.28 (3.49)	5.30 (3.79)	0.888
Vitamin B6 (mg)	1.60 (0.61)	1.61 (0.70)	0.343
Vitamin D (µg)	2.81 (1.80)	2.90 (2.10)	0.529
Folate (µg)	266.83 (104.86)	271.40 (119.96)	0.281
Energy (kcal)	1890.53 (635.34)	1864.79 (632.96)	0.228

Similar to the non-parametric tests above, when analysed in a logistic model adjusting for relevant confounders, daily intakes of alcohol were not significantly associated with global DNA methylation ($OR_{Q4\ vs.\ Q1}$: 1.05, 95%CI: 0.57 – 1.92) (Table 14). Inclusion of an interaction term between folate and alcohol in the model did not yield any significant results.

However, contrary to the non-parametric tests above that do not account for confounding, in a logistic regression model adjusting for relevant confounders, highest folate intake was significantly associated with lower probability of hypomethylation (as defined in this thesis) when compared to lowest folate intake ($OR_{Q4\ vs.\ Q1}$: 0.37, 95%CI: 0.18 – 0.76) (Table 14). Inclusion of an interaction term between folate and alcohol in the model did not alter the results.

Lastly, highest daily caloric intake significantly increased risk of hypomethylation ($OR_{Q4\ vs.\ Q1}$: 2.50, 95%CI: 1.20 – 5.18) in a logistic regression model adjusting for centre, plate, case/control status, age at blood collection, smoking status, BMI, physical activity, education, and folate and alcohol intakes (Table 14).

Table 14 – Associations between alcohol, folate and energy daily intakes with percent global methylation in a logistic regression model adjusting for relevant confounders

Daily Intake	N (%)	Mean Methylation (SD)	Odds Ratio (OR)[¤]	95% Confidence Interval
Alcohol				
Q1 = ≤0.24 g/day	91 (24.2)	70.8 (6.3)	1.00	
Q2 = >0.24 - ≤2.69 g/day	98 (26.1)	70.9 (5.7)	0.72	0.39 – 1.32
Q3 = >2.69- ≤11.18 g/day	87 (23.1)	69.9 (6.3)	1.15	0.62 – 2.13
Q4 = >11.18 g/day	100 (26.6)	70.8 (4.7)	1.05	0.57 – 1.92
Folate				
Q1 = ≤225.4 µg/day	99 (26.3)	70.2 (5.4)	1.00	
Q2 = >225.4 - ≤268.5 µg/day	89 (23.7)	71.5 (4.0)	0.57	0.31 – 1.06
Q3 = >268.5 - ≤335.8 µg/day	96 (25.5)	69.9 (7.1)	0.76	0.40 – 1.44
Q4 = >335.8 µg/day	92 (24.5)	70.9 (5.9)	0.37	0.18 – 0.76
Calories				
Q1 = ≤1582.529 /day	94 (25.0)	71.2 (5.0)	1.00	
Q2 = >1582.529 - ≤1876.68 /day	94 (25.0)	70.2 (5.3)	1.73	0.93 – 3.20
Q3 = >1876.68 - ≤2215.177 /day	94 (25.0)	70.9 (5.1)	1.64	0.85 – 3.14
Q4 = >2215.177 /day	94 (25.0)	70.0 (6.4)	2.50	1.20 – 5.18

[¤] OR in a logistic regression model with above and below median methylation as the outcome, adjusting for centre, plate, case/control status, age at blood collection, smoking status, BMI, physical activity, total caloric intake and education. For Caloric intake, OR was adjusted for centre, plate, case/control status, age at blood collection, smoking status, BMI, physical activity, education, and folate and alcohol intakes.

Given that no food group was differentially consumed between the two methylation groups, it was not surprising that adherence to the Mediterranean diet, measured with a score based on food group intakes, was also not associated with percent global methylation whether treated as a continuous ($OR_{unit\ increase\ in\ score}$: 0.98, 95%CI: 0.86 – 1.13) or as a categorical variable ($OR_{score7\text{-}9\ vs.\ score0\text{-}3}$: 0.75, 95%CI: 0.33 – 1.73) (Table 15).

Table 15 – Association between adherence to the Mediterranean diet and above/below median global methylation in a logistic regression model

Adherence to Mediterranean Diet - Score	N (%)	Mean Methylation (SD)	Odds Ratio (OR) [¤]	95% Confidence Interval
Score – Continuous			0.98	0.86 – 1.13
Score – Categorical				
0-3	139 (37.0)	70.3 (6.6)	1.00	
4-6	202 (53.7)	70.8 (5.2)	0.80	0.49 – 1.31
7-9	35 (9.3)	70.7 (5.2)	0.75	0.33 – 1.73

[¤] OR in a logistic regression model with above and below median methylation as the outcome, adjusting for centre, plate, case/control status, age at blood collection, smoking status, BMI, total caloric intake, physical activity and education.

In contrast to the Mediterranean diet score, the dietary methyl intake score did evidence a significant association with percent global methylation (Table 16). When compared to a score of 0, any non-zero value of the methyl intake score was associated with decreased likelihood of hypomethylation. This association remained even after

adjustment for relevant confounders such as centre, plate number, age at blood collection, smoking status, BMI, total caloric intake, physical activity and education. It is important to note here the small number of subjects in the reference category (score=0) and the fact that these subjects - with below median intakes of folate and vitamins B1, B2, B6 and B12 in combination with above median consumption of alcohol - are most likely severely deprived of methyl groups.

Table 16 – Association between dietary methyl intake and above/below median global methylation in a logistic regression model

Dietary Methyl Intake - Score	N (%)	Mean Methylation (SD)	Odds Ratio (OR)[¤]	95% Confidence Interval
Score – Continuous			0.86	0.84 – 0.99
Score – Categorical				
0	28 (7.4)	67.6 (6.0)	1.00	
1-2	130 (34.6)	71.2 (4.6)	0.16	0.05 – 0.46
3-4	119 (31.7)	71.0 (5.9)	0.14	0.05 – 0.43
5-6	99 (26.3)	70.2 (6.6)	0.14	0.04 – 0.44

[¤] OR in a logistic regression model with above and below median methylation as the outcome, adjusting for centre, plate, case/control status, age at blood collection, smoking status, BMI, total caloric intake, physical activity and education.

3.1.1.3 Anthropometry

Entering both height and weight as categorical variables in a logistic regression model, adjusting for relevant confounders, did not produce any significant associations. Neither height nor weight were statistically significantly associated with percent global methylation (Height: $OR_{Q4\ vs.\ Q1}$: 0.75 95%CI: 0.39 – 1.45, Weight: $OR_{Q4\ vs.\ Q1}$: 0.98 95%CI: 0.82 – 1.85).

However, in this model the regression constant was high because of the significant correlation between height and weight, even when treated as categories. Therefore two separate models were built, one examining BMI and the second examining height. BMI was treated as categorical variable (normal: ≤25kg/m^2, overweight: >25 and ≤30 kg/m^2, obese: >30kg/m^2) and this model was adjusted for centre, plate number, case-control status, age at blood collection, smoking status, total caloric energy intake, percent energy intake from fat and sex adjusted total physical activity. BMI was not significantly associated with percent global methylation in this model ($OR_{\geq 30\ kg/m2\ vs.\ <25\ kg/m2}$: 1.31, 95%CI: 0.69 – 2.50) (Table 17). Height was also treated as a categorical variable – in quartiles – and the model was adjusted for centre, plate number, case-control status, age at blood collection, smoking status and total caloric energy intake. In this model, subjects with height >161 - ≤167.5 cm vs. subjects with height ≤156 cm had a statistically significant increased risk of hypomethylation (OR: 2.20, 95% CI: 1.20-4.06) but the increased risk was not observed for subjects with height >167.5cm (OR: 0.75, 95% CI: 0.39-1.43) (Table 17).

Similarly, neither waist-to-hip ratio nor waist circumference were significantly associated with global methylation in logistic regression models adjusting for relevant confounders (Table 17).

Table 17 – Anthropometric variables and association with global methylation in a logistic regression model

Anthropometric Variable	N (%)	Mean Methylation (SD)	Odds Ratio (OR)	95% Confidence Interval
BMI – Categorical[¤]				
<25 kg/m^2	182 (48.4)	70.6 (6.1)	1.00	
25-30 kg/m^2	141 (37.5)	70.8 (5.1)	1.08	0.68 – 1.71
≥30 kg/m^2	53 (14.1)	70.1 (6.1)	1.31	0.69 – 2.50
Height – Categorical [≠]				
Q1 = ≤155.5 cm	90 (23.9)	71.3 (4.1)	1.00	
Q2 = >155.5 - ≤160 cm	105 (27.9)	70.2 (5.8)	1.20	0.67 – 2.12
Q3 = >160 - ≤165 cm	96 (25.5)	69.3 (7.4)	2.20	1.20 – 4.06
Q4 = >165 cm	85 (22.6)	71.9 (4.6)	0.75	0.39 – 1.43
Waist-to-hip ratio – Categorical [¤]				
Q1 = ≤0.75	111 (29.5)	70.5 (6.5)	1.00	
Q2 = >0.75 - ≤0.79	75 (20.0)	70.2 (5.5)	0.72	0.40 – 1.32
Q3 = >0.79 - ≤0.84	115 (30.5)	70.7 (5.9)	0.77	0.45 – 1.32
Q4 = >0.84	75 (20.0)	71.1 (4.4)	0.74	0.40 – 1.38
Waist Circumference – Categorical [¤]				
Q1 = ≤73.25 cm	94 (25.0)	79.9 (6.6)	1.00	
Q2 = >73.25 - ≤80.00 cm	98 (26.1)	71.1 (5.3)	0.81	0.45 – 1.45
Q3 = >80.00 - ≤87.65 cm	90 (23.9)	70.8 (5.6)	0.68	0.37 – 1.26
Q4 = >87.65 cm	94 (25.0)	70.6 (5.3)	0.86	0.46 – 1.60

[¤] Adjusted for centre, plate number, case-control status, age at blood collection, smoking status, total caloric energy intake, percent energy intake from fat and sex adjusted total physical activity

≠ Adjusted for centre, plate number, case-control status, age at blood collection, smoking status and total caloric energy intake

3.1.1.4 Age at Menarche – Menopausal Status – Age at Menopause – ELEE

From separate regression models, it became evident that age at menarche – whether treated as continuous or categorical – was significantly associated with global methylation (Table 18). For each year later that menarche occurred, there was a 32% increased risk of hypomethylation ($OR_{/year}$: 1.32 95%CI: 1.15 – 1.53). When treated as a categorical variable, ages at menarche 12-14 and ≥15, increased risk of hypomethylation by 76 and 354% respectively when compared to an age at menarche of ≤11 years ($OR_{12\text{-}14\ years\ vs.\ \leq 11\ years}$: 1.76 95%CI: 1.00 – 3.11 and $OR_{\geq 15 years\ vs.\ \leq 11\ years}$: 4.54 95%CI: 2.02 – 10.19). On the other hand, menopausal status, age at menopause and ELEE were not significantly associated with percent global methylation, whether treated as continuous or categorical variables (Menopausal status: $OR_{postmenopausal\ vs.\ premenopausal}$: 1.99 95%CI: 0.92 – 4.29, Age at menopause: $OR_{>50years\ vs.\ \leq 50years}$: 0.73 95%CI: 0.39 – 1.35, ELEE: $OR_{Q4\ vs.\ Q1}$: 0.75 95%CI: 0.39 – 1.43) (Table 18). In a model entering both age at menarche and ELEE, age at menarche retained its significance. ELEE and age at menopause could not be examined in the same model due to high collinearity between the two variables (r=0.9495, p<0.0001).

Table 18 – Reproductive variables and association with global methylation in a logistic regression model

Reproductive Variable	N (%)	Mean Methylation (SD)	Odds Ratio (OR)	95% Confidence Interval
Age at Menarche – Continuous ¤			1.32	1.15 – 1.53
Age at Menarche – Categorical ¤				
≤11 years	72 (19.4)	71.5 (4.5)	1.00	
12-14 years	242 (65.2)	70.5 (5.9)	1.76	1.00 – 3.11
≥15 years	57 (15.4)	69.5 (6.3)	4.54	2.02 – 10.19
Menopausal Status – Categorical≠				
Premenopausal	175	70.9 (5.2)	1.00	
Postmenopausal	201	70.4 (6.2)	1.99	0.92 – 4.29
Age at Menopause – Continuous≠			1.05	0.99 – 1.11
Age at Menopause – Categorical ≠				
≤50 years	123 (63.7)	70.2 (5.8)	1.00	
>50 years	70 (36.3)	70.7 (6.9)	0.73	0.39 – 1.35
ELEE¥ **– Continuous** ¤			1.01	0.96 – 1.06
ELEE – Categorical ¤				
Q1 = ≤31 years	120 (32.9)	71.1 (4.7)	1.00	
Q2 = >31 - ≤34 years	78 (21.4)	70.2 (6.3)	1.00	0.55 – 1.83
Q3 = >34 - ≤37years	79 (21.6)	69.9 (5.3)	1.50	0.80 – 2.80
Q4 = >37 years	88 (24.1)	70.7 (6.9)	0.75	0.39 – 1.43

¤ Adjusted for centre, plate number, case-control status, age at blood collection, smoking status, height and weight

≠ Adjusted for centre, plate number, case-control status, age at blood collection and smoking status

¥ ELEE = Estimated Lifetime Oestrogen Exposure

3.2 Genome Wide Methylation (Illumina 450K BeadChip)

3.2.1 Normalisation and Pre-processing

After exclusion of all probes which failed in more than 20% of samples, 444,278 probes out of the 485,577 remained. In addition, three subjects – all female – were excluded because they contained less than 95% detected probe readings. After these subjects' exclusion, the complete dataset consisted of 667 subjects with reliable methylation measurements and the dataset which included only females consisted of 485 subjects.

Demographic, anthropometric, lifestyle, and reproductive characteristics for these remaining subjects are presented in Table 19.

Table 19 – Demographic, anthropometric, lifestyle, and reproductive characteristics for subjects with Genome-Wide Methylation measurements

Covariate			No. Subjects with available information
Gender			
- **Male**	n (%)	182 (27.3)	
- **Female**	n (%)	485 (72.7)	667
Age	Range	34 - 72	
	Median	55	
	Mean (sd)	53.6 (7.0)	667
Birth during the WWII:			
- **No**	n (%)	461 (69.1)	
- **Yes**	n (%)	206 (30.9)	667
Highest Educational Level			
- **University**	n (%)	60 (9.2)	
- **Secondary School**	n (%)	143 (22.0)	
- **Technical/Professional school**	n (%)	86 (13.2)	
- **Primary School or no education**	n (%)	361 (55.6)	650
Height	Range	137.5 – 187.8	
	Median	161	
	Mean (sd)	162.0 (8.3)	667
Weight	Range	42.8 – 115.0	
	Median	65.8	
	Mean (sd)	67.8 (12.6)	667
BMI[±]			
- **<25 kg/m^2**	n (%)	311 (46.6)	
- **25-30 kg/m^2**	n (%)	257 (38.5)	
- **≥30 kg/m^2**	n (%)	99 (14.9)	667
Waist-to-hip ratio	Range	0.6 – 1.3	
	Median	0.8	
	Mean (sd)	0.8 (0.1)	667
Waist Circumference	Range	61 – 119.5	
	Median	83.0	
	Mean (sd)	83.9(11.8)	667
Physical Activity (hrs/week)	Range	0 – 29.5	

	Median	8.5	
	Mean (sd)	8.9 (4.5)	667
Smoking Status			
- **Current Smoker**	n (%)	142 (21.8)	
- **Former Smoker**	n (%)	169 (26.0)	
- **Never**	n (%)	340 (52.2)	651
Daily energy intake (calories/day)	Range	506.5 – 5465.9	
	Median	2084.1	
	Mean (sd)	2162.7 (641.4)	650
Daily alcohol consumption (g/day)	Range	0 – 118.8	
	Median	6.6	
	Mean (sd)	14.2 (18.8)	650
Daily folate consumption (µg/day)	Range	45.3 – 771.6	
	Median	236.4	
	Mean (sd)	249.5 (83.2)	650
Italian Mediterranean Diet Index	Range	0 – 10	
	Median	4	
	Mean (sd)	4.0 (2.0)	649
Methyl Intake Score	Range	0 – 6.0	
	Median	3.0	
	Mean (sd)	3.0 (2.0)	650
Age at Menarche			
- **≤ 11 yrs**	n (%)	80 (16.8)	
- **12 – 14 yrs**	n (%)	346 (72.7)	
- **≥ 15 yrs**	n (%)	50 (10.5)	476
Age at FFTP[±]			
- **<25 yrs**	n (%)	172 (41.2)	
- **25 – 30 yrs**	n (%)	171 (40.9)	
- **≥30 yrs**	n (%)	75 (17.9)	418
Parous			
- **No**	n (%)	17 (3.9)	
- **Yes**	n (%)	418 (96.1)	435
Breastfeeding			
- **No**	n(%)	169 (34.8)	
- **Yes**	n(%)	316 (65.2)	485
Menopausal Status			
- **Premenopausal**	n (%)	172 (35.9)	
- **Postmenopausal**	n (%)	307 (64.1)	479
Age at menopause			
- **≤50 years**	n (%)	192 (64.4)	
- **>50 years**	n (%)	106 (35.6)	298
ELEE	Range	20 - 52	
	Median	36	
	Mean (sd)	35.1 (5.1)	476
HRT[±] **use**			
- **Ever**	n (%)	80 (16.8)	
- **Never**	n (%)	395 (83.2)	475
OC[±] **use**			
- **Ever**	n (%)	185 (38.8)	
- **Never**	n (%)	292 (61.2)	477

± BMI: Body Mass Index, FFTP: First Full Term Pregnancy, HRT: Hormone Replacement Therapy, OC: Oral Contraceptive

The complete dataset successfully underwent background subtraction and dye bias correction as shown in Figure 18.

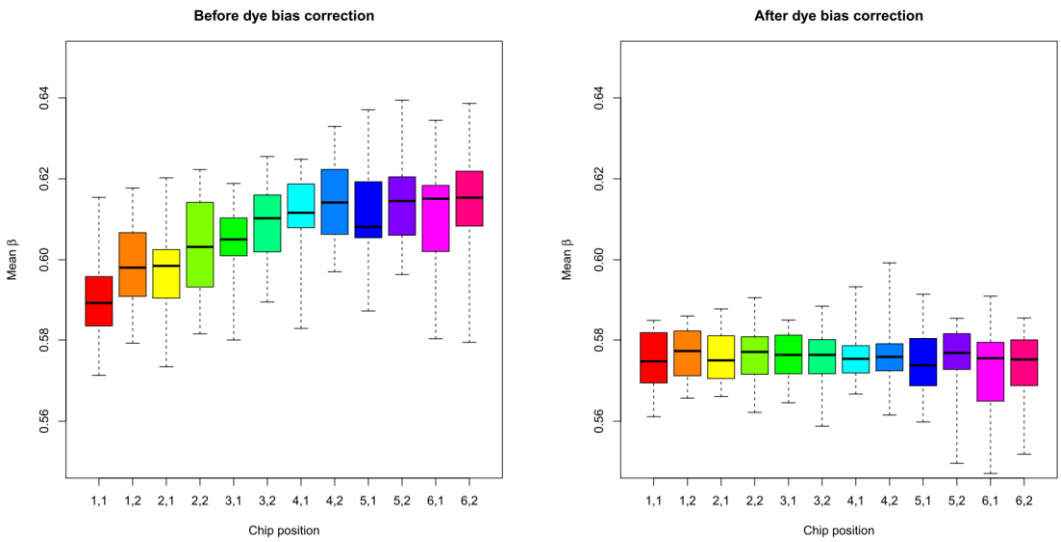


Figure 18 – Diagnostic plot for dye bias correction following background subtraction.

Following background subtraction and dye-bias correction, the mean and interquartile range of methylation β-values for both colour intensities were standardized. Figure kindly provided by Gianluca Campanella.

In addition, BMIQ successfully normalised type I and type II probe methylation values. As indicatively shown for one sample (subject #100) in Figure 19, BMIQ moved the type 2-BMIQ peaks towards the extremes of the distribution and more in line with the peaks of type 1 probes. In addition, the areas under the type 2-BMIQ curve at methylation β-values of <20 and >80 were the same as the respective areas under the type 2 curve. This indicates that, after BMIQ normalization, type II probes retained their proportions of unmethylated, hemi-methylated and methylated probes, as expected.

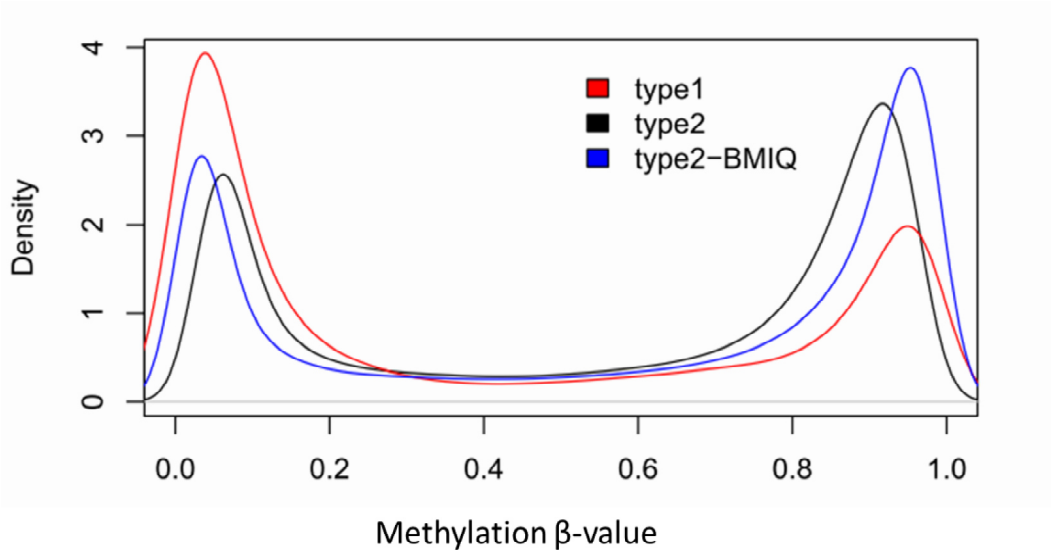


Figure 19 – Plot showing BMIQ normalization of type 1 (type I) and type 2 (type II) probes for a random sample (subject #100).

The x-axis denotes methylation β-values and the y-axis the density of readings at each β-value. Three curves are shown: red – type I probes, black – type II probes before normalization, and blue – type II probes after BMIQ normalization.

Lastly, ComBat successfully removed batch effects by normalising the median and standard deviation of the methylation of all probes in each chip as presented in Figure 20 for the entire dataset and in Figure 21 for the "females only" dataset.

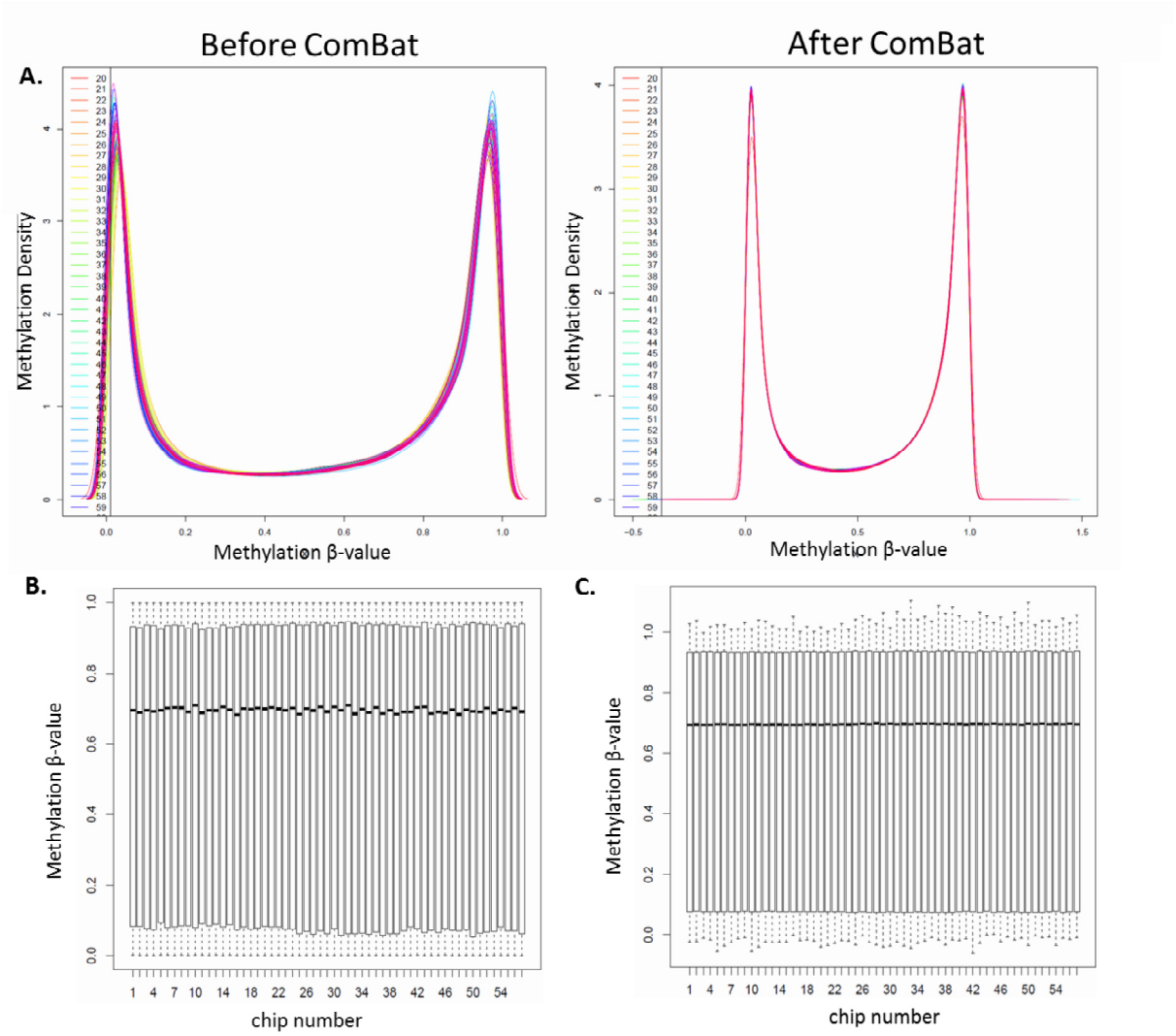


Figure 20 – Removal of batch effects in the entire dataset (n=667, 57 chips) using ComBat.

Panel A: shows the methylation densities of all probes in each chip before (left) and after (right) Combat. Each chip is represented by a differently coloured line. Methylation densities converge after ComBat. Panel B: Boxplots of median and interquartile β-value methylation of each chip before Combat. Panel C: Boxplots of median and interquartile methylation of each chip after Combat – medians and interquartile ranges were standardised across chips.

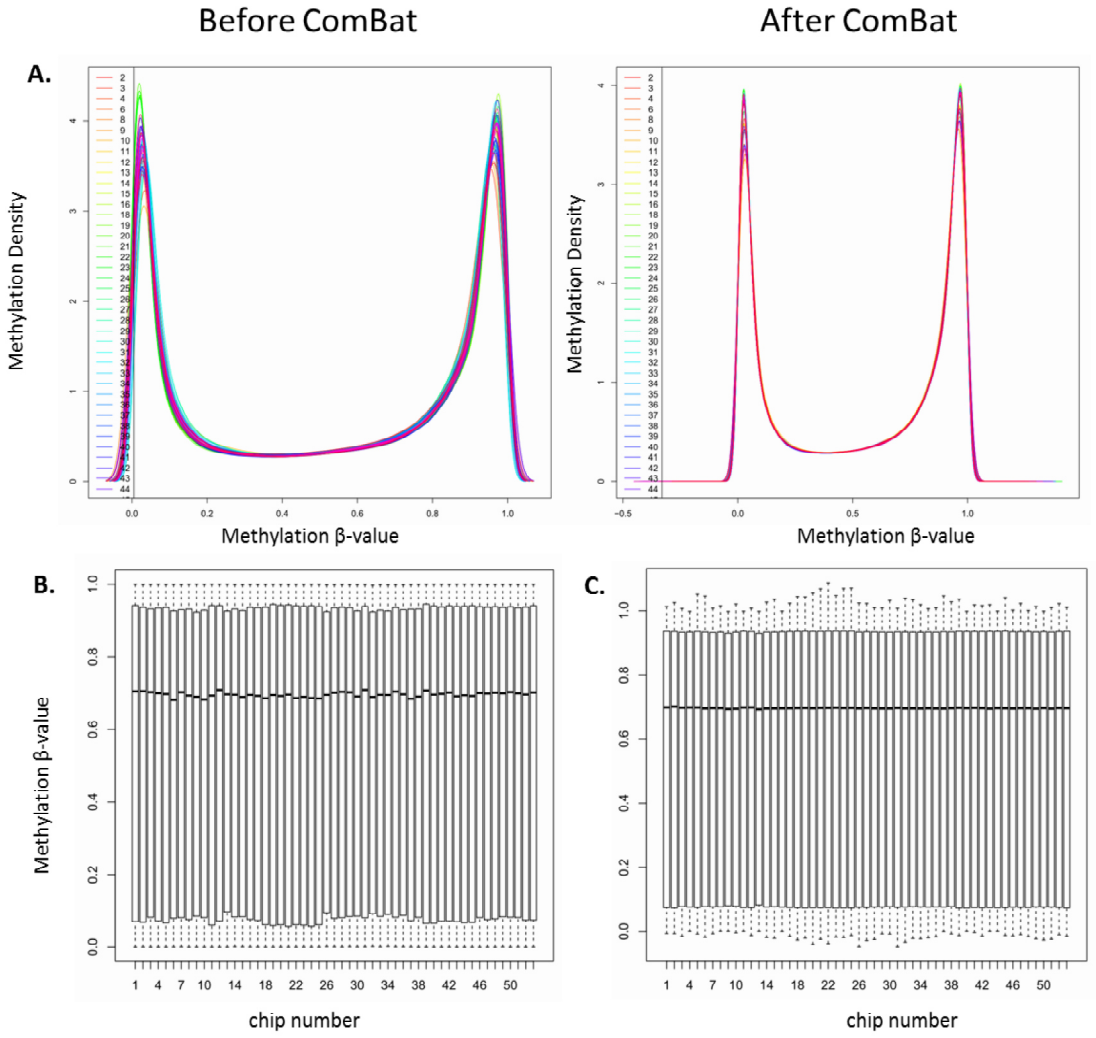


Figure 21 – Removal of batch effects in the "females only" dataset (n=485, 53 chips) using ComBat.

Panel A: shows the methylation densities of all probes in each chip before (left) and after (right) Combat. Each chip is represented by a differently coloured line. Methylation densities converge after ComBat. Panel B: Boxplots of median and interquartile β-value methylation of each chip before Combat. Panel C: Boxplots of median and interquartile methylation of each chip after Combat – medians and interquartile ranges were standardised across chips.

3.2.2 Whole Genome Methylation analysis

Following pre-processing and normalisation and before whole genome methylation and clustering analyses, several additional probes were removed because of their imperfect design, according to Price et al.[398], or because they were sex chromosome loci that could lead to gender bias, or because they were found to be differentially methylated in WBCs. Table 20 shows the number of probes retained for whole genome methylation and clustering analyses, following exclusion of these probes. Figure 22 summarizes the genomic environments of these probes, classifying them according to functional genomic distribution (FGD), CpG island context and chromosomal position.

Table 20 – Number of probes retained for whole genome methylation and clustering analyses, following removal of compromising probes

Step	No. of Probes retained
Initial	485,577
Exclusion of probes with >20% failed samples (for all analyses)	444,278
Exclusion of probes with at least 1 SNP in their probe body	362,516
Exclusion of probes with at least 1 SNP at the target CpG	357,292
Exclusion of autosomal probes cross-hybridizing to XY loci/chromosomes	348,216
Exclusion of XY probes cross-hybridizing to autosomal loci/chromosomes	327,587
Exclusion of probes whose entire body mapped to repetitive elements	309,751
Exclusion of probes located on X or Y chromosomes	301,822
Exclusion of probes differentially methylated in WBCs	293,873

All 293,873 probes were combined for each subject to get an estimate for whole genome methylation, by calculating the median. In addition, probes in each FGD class and CpG Island context class were combined to get median probe category specific methylation measurements for each individual. Independently, 32,873 probes were found to match to repetitive elements of the genome and were combined to get an estimate of repetitive element methylation. The results of the investigation on whether these probe category specific methylation values were significantly different between subjects in different exposure categories are presented in Table 21. Table 21 only presents P-values from analyses on all subjects ($n_{\text{both sexes}}$=667 and $n_{\text{females only}}$=485) because results from analyses on only those subjects that remained healthy for at least five years following blood collection ($n_{\text{both sexes}}$=504 and $n_{\text{females only}}$=357) did not differ significantly.

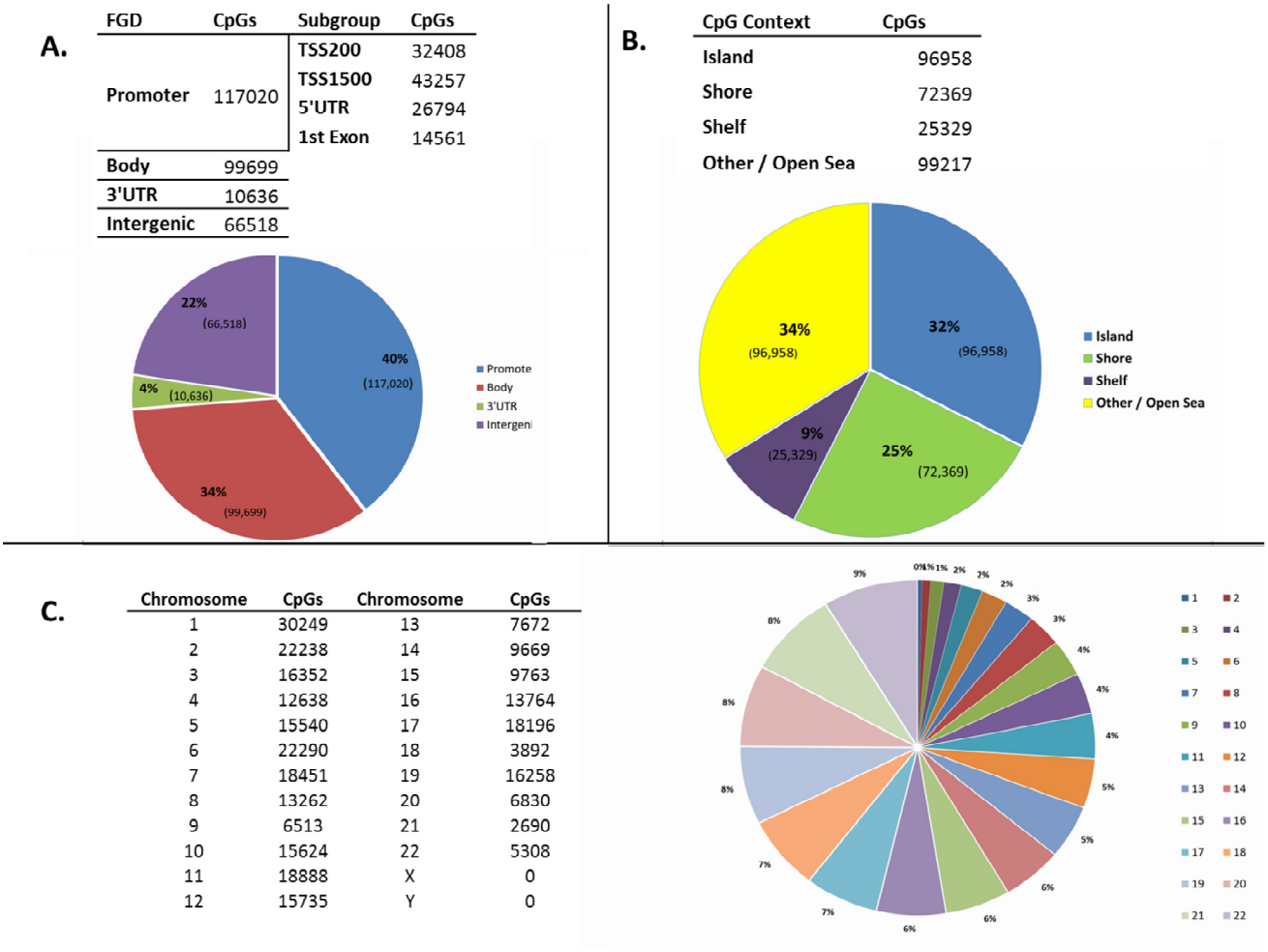


Figure 22 – Description of the 293,873 retained probes for whole genome methylation analyses.

A) Number of probes classified according to functional genomic distribution (FGD): promoter, body, 3'UTR, Intergenic. B) Number of probes classified according to CpG Island context: island, shore, shelf, other. C) Number of probes classified according to chromosomal position.

To replicate the results observed with LUMA, median methylation across all probes per subject was examined between different exposure categories. In addition, to investigate which probes contribute most to the difference in methylation between exposure categories, median methylation across different groups of probes based on location or function were examined against the exposures of interest. Reassuringly, as expected from previous reports [408], CpG island shores were found to be among the most variable genomic regions .

As shown in Table 21, whole genome methylation (all probes) was significantly different between highest combined household occupational status, caloric intake, folate intake, and dietary methyl intake categories (P-values: 0.047, 0.020, 0.022, and 0.026 respectively). The difference in whole genome methylation between subjects with different highest combined household occupational position is mostly attributed to differential gene promoter, island and shore probe methylation (P-values: 0.018, 0.016 and 0.029 respectively). Similarly, caloric intake and folate intake differential whole genome methylation was mostly explained by methylation changes in shore probes (P-values $_{Calories}$: 0.023 respectively & P-values $_{Folate}$: 0.037). No single probe category could explain

the difference in whole genome methylation between different dietary methyl intake categories, suggesting that methylation changes are spread throughout the genome and not localized. On the other hand, even though whole genome methylation did not differ, CpG Island shore methylation was significantly different between subjects in the four educational categories (P-value: 0.037). Similarly, subjects in different household density categories significantly differed in their gene promoter and island methylation (P-values: 0.029 and 0.044, respectively). However, given the multiple associations examined, the Bonferroni-corrected threshold of significance ($P<0.0003$) was not exceeded by any of the above mentioned associations (Table 21).

Table 21 – Associations between each exposure variable and average methylation across different functional genomic distribution and CpG Island context probe classes.

P-values are from linear regression tests with methylation as a continuous outcome and the exposure as a categorical variable, adjusting for relevant confounders as outlined in Table 8. Yellow boxes highlight significant associations (P<0.05) whereas Blue boxes highlight associations that remain significant even after Bonferroni correction for multiple testing ($P_{Bonferroni} < 0.0003$).

	Exposure Variable																	
	Education (4 categories)	Education (2 categories)	Household Density	Household's highest occupational position	Father's occupational position	When household first had fridge	Birth during WWII	Italian Mediterranean Diet index	Total Caloric Intake	Alcohol Intake	Folate Intake	Methyl Intake Score	BMI	Height	Age at menarche	Menopausal Status	Age at menopause	Estimated Lifetime Oestrogen Exposure
Average Methylation across:																		
All probes	0.064	0.449	0.741	0.047	0.116	0.614	0.602	0.215	0.020	0.677	0.022	0.026	0.734	0.933	0.321	0.138	0.593	0.447
Gene promoter probes	0.240	0.199	0.029	0.018	0.996	0.635	0.259	0.418	0.522	0.835	0.482	0.768	0.391	0.833	0.755	0.652	0.176	0.687
Gene body probes	0.717	0.999	0.179	0.153	0.559	0.333	0.251	0.937	0.378	0.861	0.452	0.176	0.968	0.720	0.954	0.465	0.204	0.444
3'UTR probes	0.689	0.590	0.715	0.153	0.426	0.313	0.254	0.998	0.120	0.854	0.240	0.186	0.349	0.897	0.565	0.230	0.360	0.509
Intergenic probes	0.447	0.625	0.066	0.627	0.402	0.690	0.317	0.928	0.274	0.812	0.390	0.138	0.630	0.315	0.671	0.607	0.154	0.485
Island probes	0.372	0.289	0.044	0.016	0.932	0.643	0.279	0.481	0.768	0.798	0.622	0.874	0.259	0.928	0.957	0.442	0.222	0.643
Shore probes	0.037	0.399	0.279	0.029	0.190	0.739	0.319	0.181	0.023	0.791	0.037	0.090	0.915	0.467	0.234	0.177	0.255	0.610
Shelf probes	0.953	0.833	0.439	0.128	0.348	0.340	0.262	0.976	0.169	0.909	0.304	0.195	0.379	0.691	0.808	0.265	0.321	0.577
Other / Open Sea probes	0.414	0.625	0.095	0.130	0.554	0.240	0.198	0.995	0.217	0.806	0.406	0.258	0.910	0.623	0.798	0.555	0.204	0.539
Repetitive element probes	0.729	0.922	0.071	0.160	0.416	0.443	0.198	0.900	0.402	0.667	0.676	0.457	0.888	0.781	0.824	0.177	0.220	0.681

3.2.3 Genome Wide locus by locus Methylation Analyses

For locus by locus analyses, after exclusion of probes which failed in more than 20% of samples, all 444,278 probes were retained. Figure 23 summarises the genomic environments of these probes, classifying them according to FGD, CpG island context and chromosomal position. The percentages of probes in each FGD class did not differ from the corresponding percentages for the 301,822 probes. However, the percentages of probes in different CpG Contexts did differ, with more open sea probes and fewer shore and island probes.

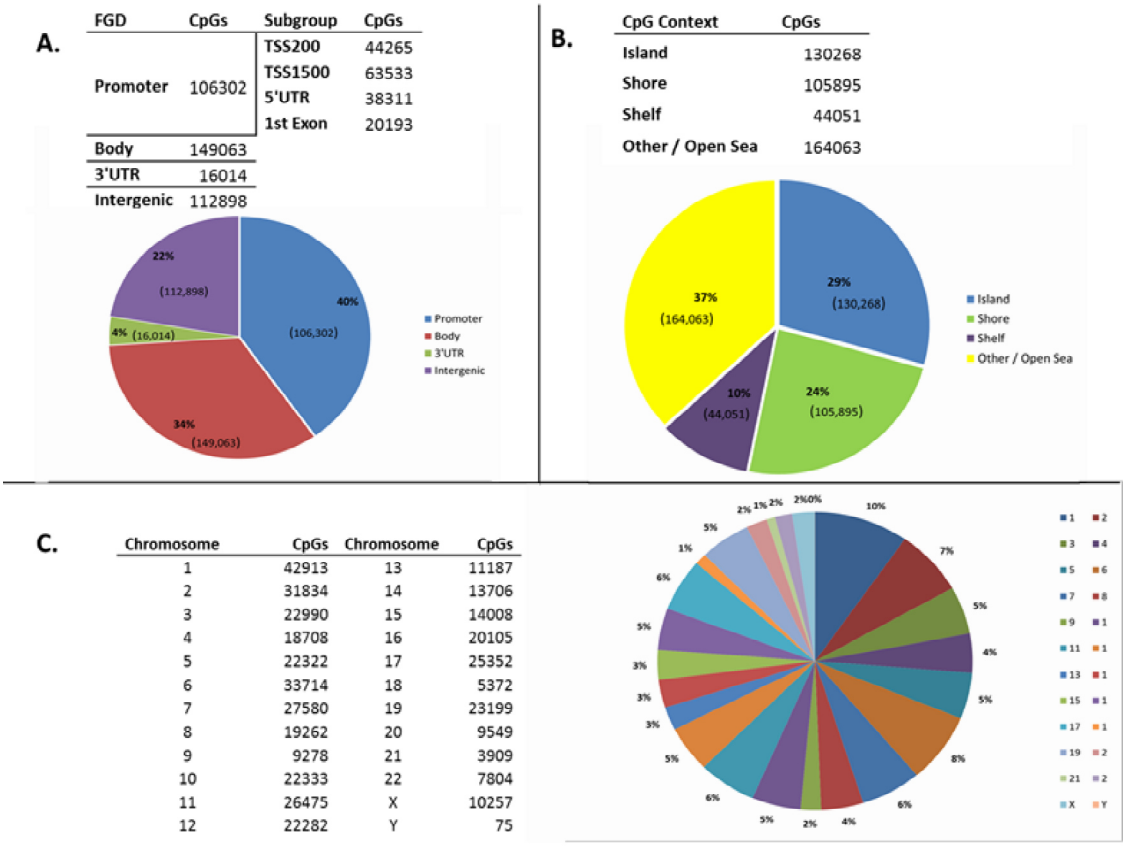


Figure 23 – Description of the 444,278 retained probes for locus by locus analyses.

A) Number of probes classified according to functional genomic distribution: promoter, body, 3'UTR, Intergenic. B) Number of probes classified according to CpG Island context: island, shore, shelf, other. C) Number of probes classified according to chromosomal position.

The results of locus by locus analyses on all subjects and on only those subjects that remained healthy for at least five years following blood collection (n=504 for analyses on both sexes and n=357 for the "females only" analyses), did not differ significantly. Therefore, the results presented below are from analyses on all subjects, unless otherwise stated. In addition, when locus-by-locus analyses were run without the probes excluded for the whole genome analyses, results did not vary. Therefore, results from analyses including all 444, 278 probes are presented.

3.2.3.1 Socioeconomic Status

3.2.3.1.1 Early life Socioeconomic Status

Exposures used as proxies for childhood SES, namely father's occupation, when household first had fridge, and birth during WWII, were not found to be significantly associated with methylation at any of the 444,278 CpG sites investigated (Figure 24).

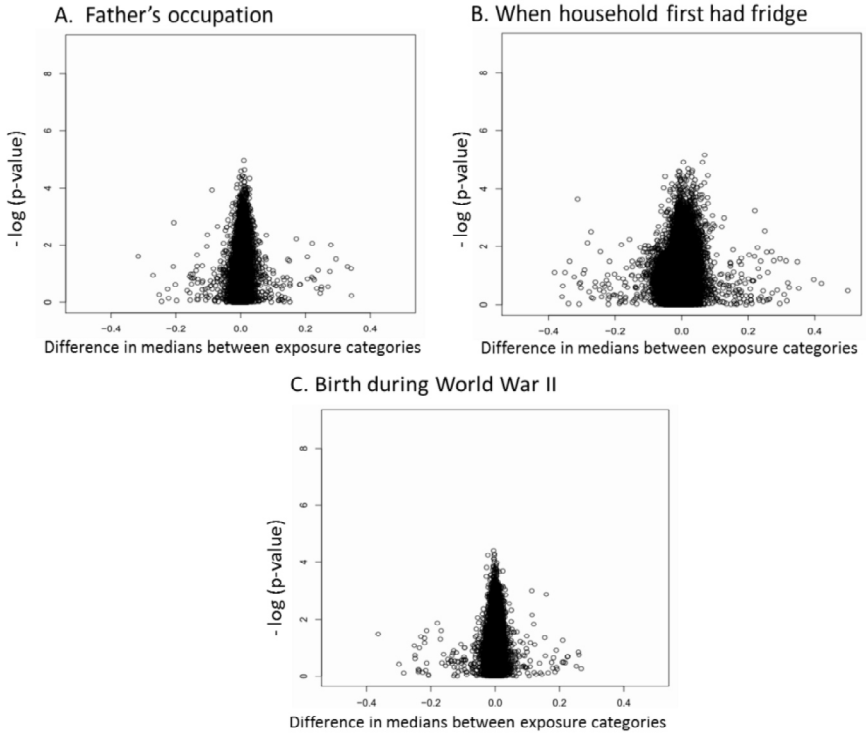


Figure 24 – Volcano plots illustrating the negative logarithm of the P-value from the regression analysis of each probe against the difference in medians between early life socioeconomic status exposure groups for that probe.

When the exposure is categorised in more than two groups, the difference in medians between the two groups at the extremes of the exposure distribution is calculated. A: Volcano plot of P-values from the linear regression analysis with methylation as a continuous outcome and father's occupation as a categorical variable (non-manual vs. manual). B: Volcano plot of P-values from the linear regression analysis with methylation as a continuous outcome and when household first had fridge as a categorical variable (>1964 vs. <1954). C: Volcano plot of P-values from the linear regression analysis with methylation as a continuous outcome and birth during WWII as a categorical variable (yes vs. no). No plot demonstrated significantly ($P \leq 1x10^{-7}$) differentially methylated CpG sites between exposure groups.

Similarly, the P-value density distributions from locus by locus analyses for each variable, did not show an enrichment of CpG sites with very small P-values indicating no enrichment of statistically significant CpGs (Figure 25). In these plots, an almost flat density line is what would be expected by chance if methylation and exposure were completely independent, whereas a negatively skewed density distribution might indicate a more diffuse but not necessarily significant effect of the exposure on methylation.

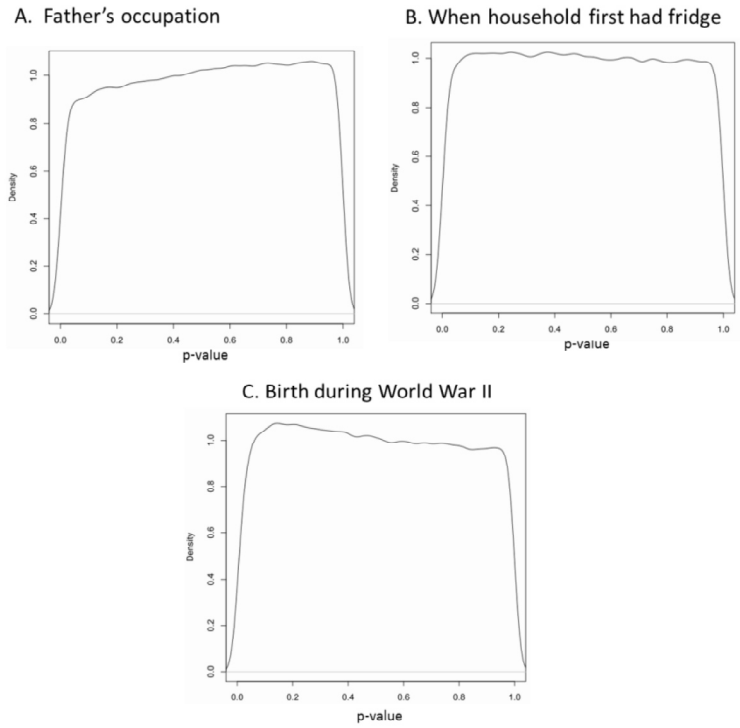


Figure 25 – Graphical representations of P-value density distributions in locus by locus analyses of early life socioeconomic status variables.

When the exposure is categorised in more than two groups, values for the comparison between the most dissimilar exposure groups are reported. An almost horizontal density line represents the uniform distribution expected by chance. A: Density distribution of P-values from the linear regression analysis with methylation as a continuous outcome and father's occupation as a categorical variable (non-manual vs. manual). B: Density distribution of P-values from the linear regression analysis with methylation as a continuous outcome and when household first had fridge as a categorical variable (>1964 vs. <1954). C: Density distribution of P-values from the linear regression analysis with methylation as a continuous outcome and birth during WWII as a categorical variable (yes vs. no).

3.2.3.1.2 Adult Socioeconomic Status

Analyses of variables indicative of adult socioeconomic circumstances, namely education, household density, and household's highest occupational position, were not significantly associated with methylation at any of the 444,278 CpG sites examined, using the strict thresholds established (Figure 26).

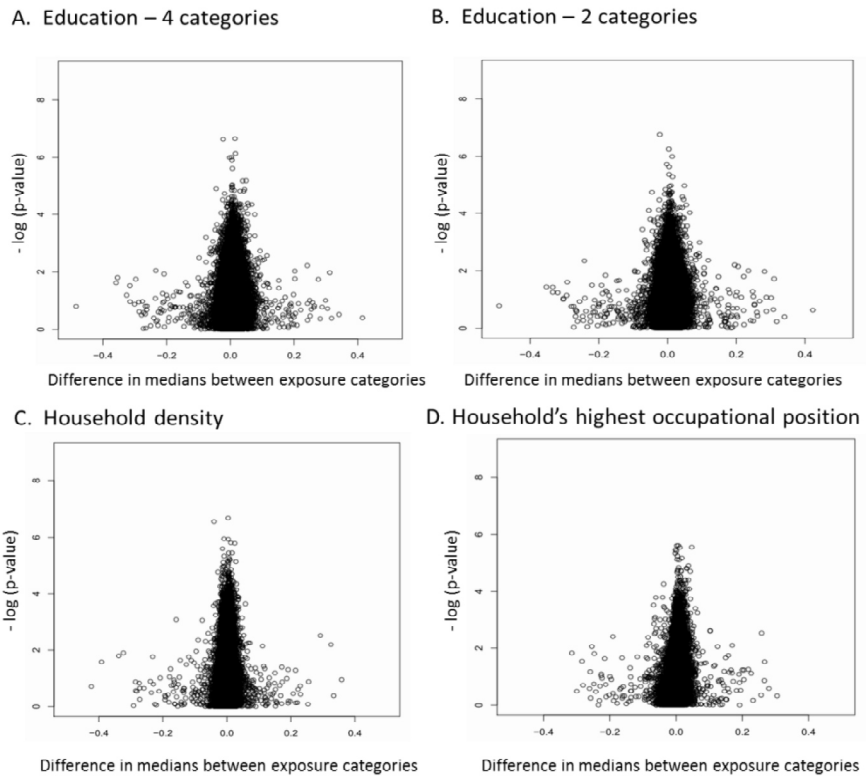


Figure 26 – Volcano plots illustrating the negative logarithm of the P-value from the regression analysis of each CpG site against the difference in medians between adult socioeconomic status exposure groups for that probe.

When the exposure is categorised in more than two groups, the difference in medians between the two groups at the extremes of the exposure distribution is calculated. A: Volcano plot of P-values from the linear regression analysis with methylation as a continuous outcome and education as a categorical variable – 4 categories (University vs. Primary school or No education). B: Volcano plot of P-values from the linear regression analysis with methylation as a continuous outcome and education as a categorical variable – 2 categories (College vs. lower than College). C: Volcano plot of P-values from the linear regression analysis with methylation as a continuous outcome and household density as a categorical variable (≥1 vs. <1 rooms/person). D: Volcano plot of P-values from the linear regression analysis with methylation as a continuous outcome and household's highest occupational position as a categorical variable (High – white collar occupation vs. Low – blue collar workers and farmers). No plot demonstrated significantly ($P \leq 1x10^{-7}$) differentially methylated CpG sites between exposure groups.

However, the skewed distributions of P-values obtained from locus by locus regression analyses, with an enrichment of CpG sites with relatively small P-values, suggest that education, whether examined in 4 categories or in 2, as well as household density may have a widespread influence on methylation (Figure 27).

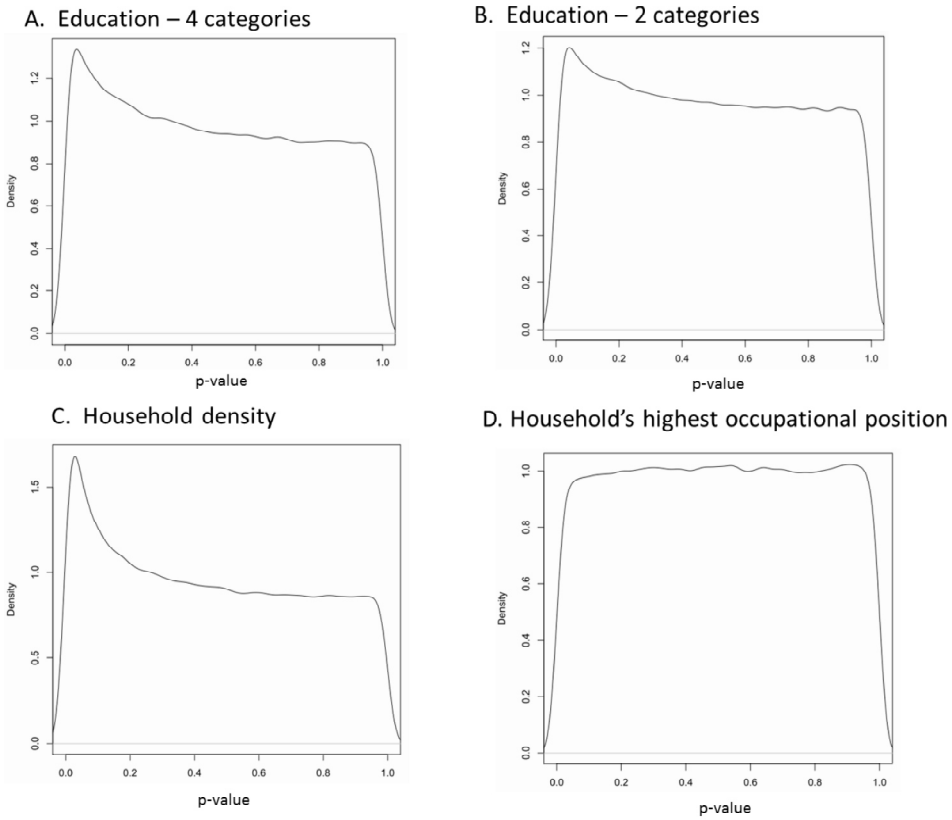


Figure 27 – Graphical representations of P-value distributions in locus by locus analyses of adult socioeconomic status variables.

When the exposure is categorised in more than two groups, values for the comparison between the most dissimilar exposure groups are reported. An almost horizontal line represents the uniform distribution expected by chance. A: Density distribution of P-values from the linear regression analysis with methylation as a continuous outcome and education as a categorical variable – 4 categories (University vs. Primary school or No education). B: Density distribution of P-values from the linear regression analysis with methylation as a continuous outcome and education as a categorical variable – 2 categories (College vs. lower than College). C: Density distribution of P-values from the linear regression analysis with methylation as a continuous outcome and household density as a categorical variable (≥1 vs. <1 rooms/person). D: Density distribution of P-values from the linear regression analysis with methylation as a continuous outcome and household's highest occupational position as a categorical variable (High – white collar occupation vs. Low – blue collar workers and farmers).

3.2.3.2 Anthropometry

None of the CpGs examined demonstrated statistically significant differential methylation between BMI or between height groups. Both anthropometric variables' analyses produced volcano plots without significant hits above the $1x10^{-7}$ threshold and the P-value distribution for height also indicated no effect on DNA methylation. However, the P-value distribution of BMI indicates an enrichment of relatively small but not quite significant P-values which may signify a more widespread effect on methylation (Figure 28).

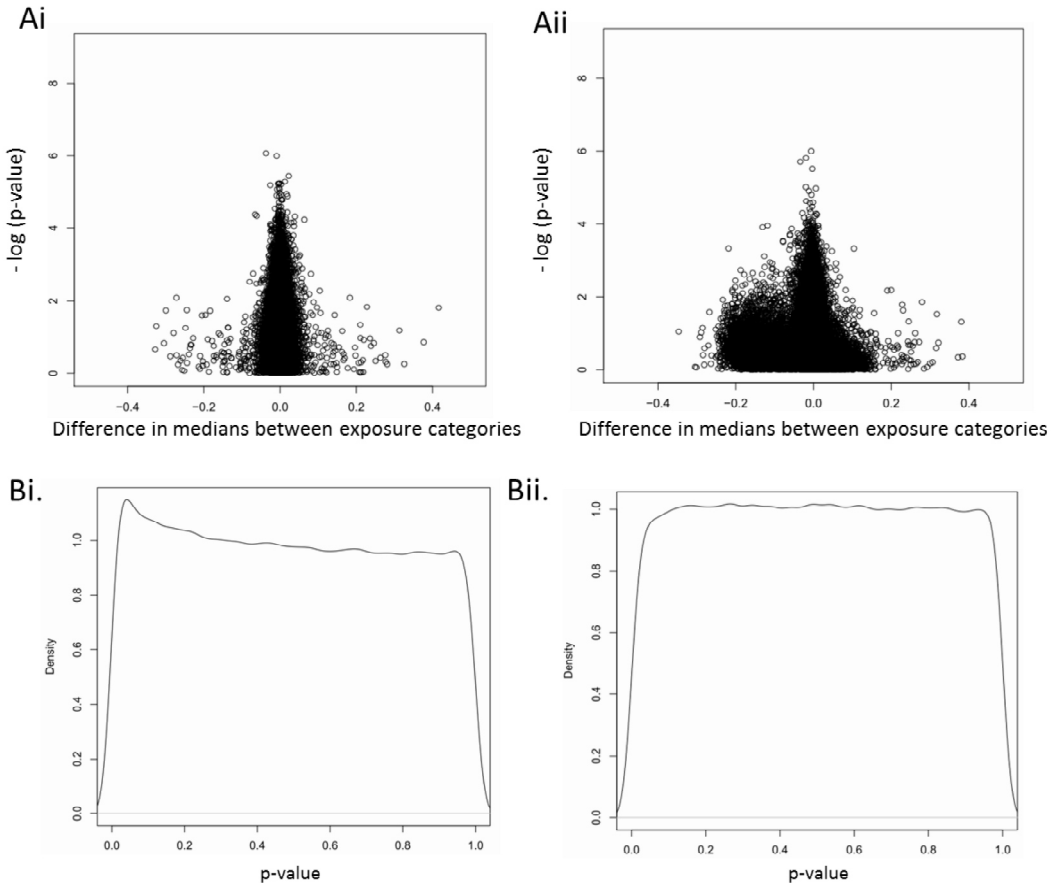


Figure 28 – A. Volcano plots illustrating the negative logarithm of the P-value from the regression analysis of each CpG site against the difference in medians between anthropometric variable exposure groups for that probe. B. Graphical representations of P-value distributions in locus by locus analyses of anthropometric variables.

When the exposure is categorised in more than two groups, the difference in medians between the two groups at the extremes of the exposure distribution is calculated. Ai: Volcano plot of P-values from the linear regression analysis with methylation as a continuous outcome and BMI as a categorical variable (≥30 vs. <25kg/m^2). Aii: Volcano plot of P-values from the linear regression analysis with methylation as a continuous outcome and height as a categorical variable (Quartile 4 vs. Quartile 1). No volcano plot demonstrated significantly (P≤1x10^{-7}) differentially methylated CpG sites between exposure groups. Bi: Density distribution of P-values from the linear regression analysis with methylation as a continuous outcome and BMI as a categorical variable (≥30 vs. <25kg/m^2) Primary school or No education). Bii: Density distribution of P-values from the linear regression analysis with methylation as a continuous outcome and height as a categorical variable (Quartile 4 vs. Quartile 1).

3.2.3.3 Nutrition

Similar to the anthropometric variables examined, none of the nutritional exposures under investigation produced significantly differentially methylated CpGs. As shown by the volcano plots in Figure 29, neither adherence to the Mediterranean Diet, nor total caloric intake, nor alcohol or folate or dietary methyl intakes significantly altered methylation at any of the 444,278 CpG loci examined.

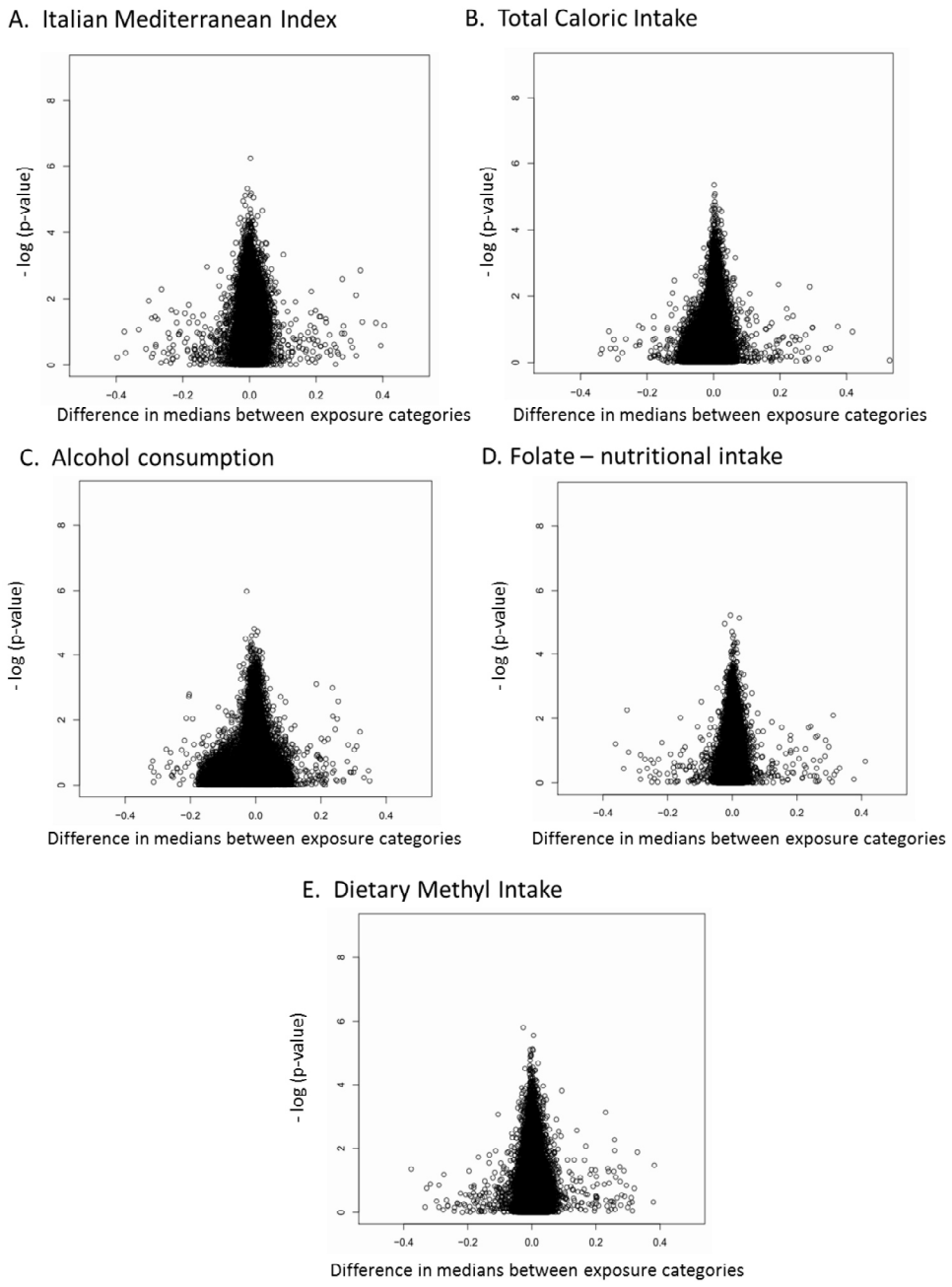


Figure 29 – Volcano plots illustrating the negative logarithm of the P-value from the regression analysis of each CpG site against the difference in medians between dietary variable exposure groups for that probe.

When the exposure is categorised in more than two groups, the difference in medians between the two groups at the extremes of the exposure distribution is calculated. A: Volcano plot of P-values from the linear regression analysis with methylation as a continuous

outcome and Italian Mediterranean Index as a categorical variable (score 7-10 vs. score 0-3). B: Volcano plot of P-values from the linear regression analysis with methylation as a continuous outcome and total caloric intake as a categorical variable (Quartile 4 vs. Quartile 1). C. Volcano plot of P-values from the linear regression analysis with methylation as a continuous outcome and alcohol consumption as a categorical variable (Quartile 4 vs. Quartile 1). D: Volcano plot of P-values from the linear regression analysis with methylation as a continuous outcome and folate intake as a categorical variable (Quartile 4 vs. Quartile 1). E: Volcano plot of P-values from the linear regression analysis with methylation as a continuous outcome and dietary methyl intake score as a categorical variable (Score 5-6 vs. Score 0). No plot demonstrated significantly ($P \leq 1x10^{-7}$) differentially methylated CpG sites between exposure groups.

The P-value distributions in Figure 30(B, C, D, E), which do not show an enrichment of relatively small P-values, support that total caloric intake, alcohol consumption, dietary folate intake, and dietary methyl intake score are not associated with DNA methylation. However, the negatively skewed distribution observed for adherence to the Mediterranean Diet, with an enrichment of CpG sites with small P-values, seems to suggest that Italian Mediterranean Index score may influence methylation (Figure 30A).

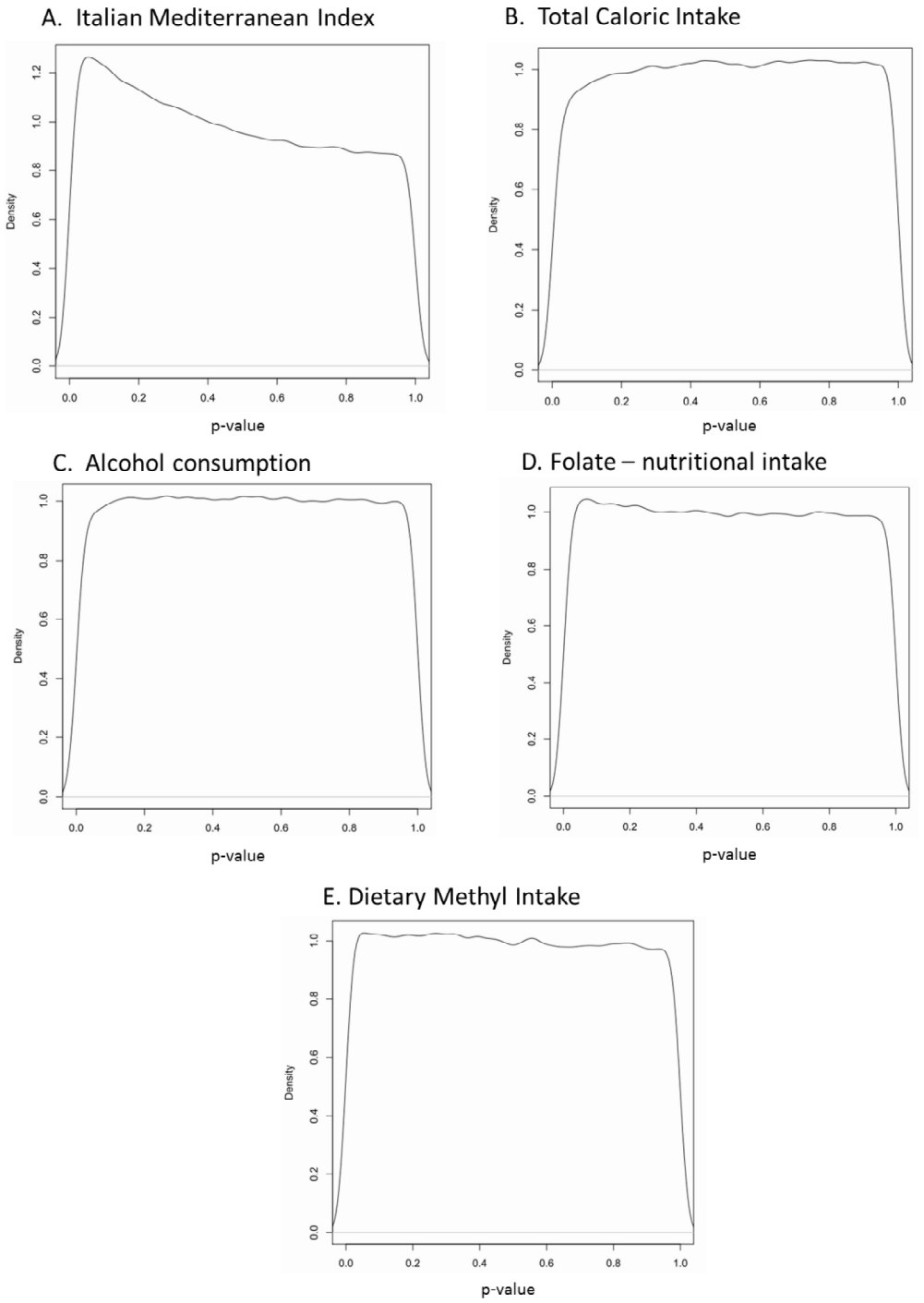


Figure 30 – Graphical representations of P-value distributions in locus by locus analyses of nutritional variables.

When the exposure is categorised in more than two groups, values for the comparison between the most dissimilar exposure groups are reported. An almost horizontal line represents the uniform distribution expected by chance. A: Density distribution of P-values from the linear regression analysis with methylation as a continuous outcome and Italian Mediterranean Index as a categorical variable (score 7-10 vs. score 0-3). B: Density distribution of P-values from the linear regression analysis with methylation as a continuous outcome and total caloric intake as a categorical variable (Quartile 4 vs. Quartile 1). C. Density distribution of P-values from the linear regression analysis with methylation as a continuous outcome and alcohol consumption as a categorical variable (Quartile 4 vs. Quartile 1). D: Density distribution of P-values from the linear regression analysis with methylation as a continuous outcome and folate intake as a categorical variable (Quartile 4 vs. Quartile 1). E: Density distribution of P-values from the linear regression analysis with methylation as a continuous outcome and dietary methyl intake score as a categorical variable (Score 5-6 vs. Score 0).

3.2.3.4 Age at Menarche – Menopausal Status – Age at Menopause – ELEE

The locus by locus analyses of the relationships between reproductive variables and probe methylation revealed only 3 significantly differentially methylated CpG sites between exposure categories (Figure 31). A menarcheal age of ≥15 vs. ≤11 years was significantly associated with methylation at two probes: cg27483455, which targets a CpG site in an Island in the body of the *ACCN2* gene and cg10903132, which targets a CpG site located in the south shelf of a CpG island. However, the probe body of cg10903132 completely matches to repetitive elements in the genome[398] therefore whether this methylation value represents the true methylation of the target CpG site is unknown. Nevertheless, for both sites, an older menarcheal age was associated with 1% increase in methylation, an effect estimate that is unlikely to have any true biological effect (Table 22). Similarly, ELEE in the highest quartile compared to ELEE in the lowest quartile was associated with a slight increase in methylation at a CpG site (cg 01781745) located in a CpG island in the 1st Exon of the *KLF9* gene (Table 22).

Not only are the differences in methylation between exposure groups too small to have a biological effect, they are also well below the 28% difference threshold which can be detected with confidence given the present sample size, significance cut off and 80% power.

It is also important to note that these significantly differentially methylated CpGs did not retain their significance when the models were run on only the females who remained healthy for the entire length of follow-up.

Table 22 – Significantly differentially methylated CpGs between age at menarche and Estimated Lifetime Oestrogen Exposure (ELEE) categories

TargetID	Median (SD)		Regression Coefficient	P-value	Q-value	Chromosome	MAPINFO	Gene	Position relative to gene	CpG Islands Name	Position relative to CpG Island
	Menarcheal Age ≤11years	**Menarcheal Age ≥15years**									
cg27483455	0.980 (0.011)	0.989 (0.011)	1.034	2.51E-09	0.001	12	50475360	ACCN2	Body	chr12:50474869-50475445	Island
cg10903132	0.964 (0.008)	0.968 (0.008)	0.333	1.01E-07	0.023	10	126081404			chr10:126077405-126078075	S_Shelf
	ELEE Q1 (≤32 years)	**ELEE Q4 (>39 years)**									
cg01781745	0.009 (0.005)	0.013 (0.006)	0.721	7.02E-08	0.031	9	73029065	KLF9	1stExon	chr9:73027330-73029184	Island

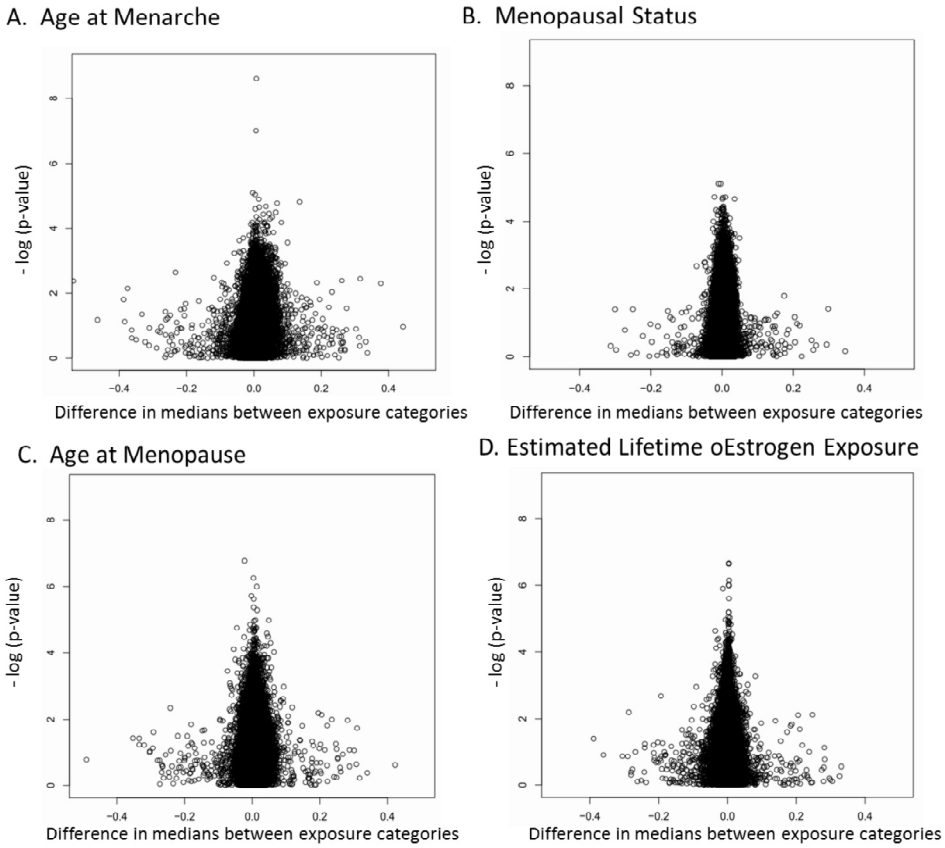


Figure 31 – Volcano plots illustrating the negative logarithm of the P-value from the regression analysis of each CpG site against the difference in medians between reproductive variable exposure groups for that probe.

When the exposure is categorised in more than two groups, the difference in medians between the two groups at the extremes of the exposure distribution is calculated. A: Volcano plot of P-values from the linear regression analysis with methylation as a continuous outcome and age at menarche as a categorical variable (≥15 vs. ≤11 years), showing two CpG sites with significant ($P \leq 1x10^{-7}$) but minor methylation differences between exposure groups. B: Volcano plot of P-values from the linear regression analysis with methylation as a continuous outcome and menopausal status as a categorical variable (post- vs. pre-menopausal). C: Volcano plot of P-values from the linear regression analysis with methylation as a continuous outcome and age at menopause as a categorical variable (>50 vs. ≤50 years). D: Volcano plot of P-values from the linear regression analysis with methylation as a continuous outcome and estimated lifetime oestrogen exposure as a categorical variable (Quartile 4 vs. Quartile 1), showing one CpG site with significant ($P \leq 1x10^{-7}$) but minor methylation differences between exposure groups.

On the other hand, the distributions of P-values obtained from locus by locus regression analyses for each reproductive variable seem to tell a different story. The skewed distributions with an enrichment of CpG sites with small P-values suggest that menopausal status, age at menopause and ELEE may be associated with diffuse DNA methylation changes, but age at menarche is not (Figure 32).

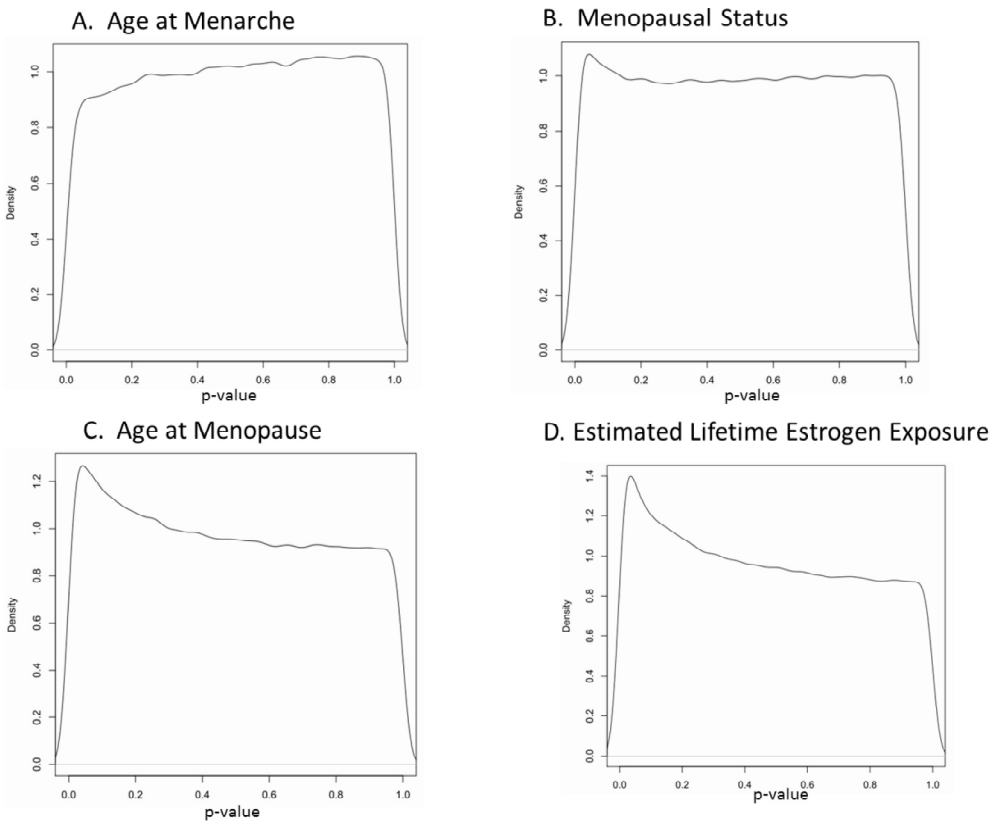


Figure 32 – Graphical representations of P-value distributions in locus by locus analyses of reproductive variables.

When the exposure is categorised in more than two groups, values for the comparison between the most dissimilar exposure groups are reported. An almost horizontal line represents the uniform distribution expected by chance. A: Density distribution of P-values from the linear regression analysis with methylation as a continuous outcome and age at menarche as a categorical variable (≥15 vs. ≤11 years). B: Density distribution of P-values from the linear regression analysis with methylation as a continuous outcome and menopausal status as a categorical variable (post- vs. pre-menopausal). C: Density distribution of P-values from the linear regression analysis with methylation as a continuous outcome and age at menopause as a categorical variable (>50 vs. ≤50 years). D: Density distribution of P-values from the linear regression analysis with methylation as a continuous outcome and estimated lifetime oestrogen exposure as a categorical variable (Quartile 4 vs. Quartile 1).

3.2.4 Clustering

Hierarchical Clustering based on Euclidean distance with average linkage failed to cluster subjects in a meaningful way, according to the exposures examined. Figure 33 shows an indicative collection of heatmaps which demonstrate that HC of the 1000 most variable probes did not cluster subjects in a way that was associated with the exposures of interest. This was evident by the lack of grouping together of similar colours on the coloured column bar. A similar observation was made for all the exposures examined, even though not all heatmaps are displayed.

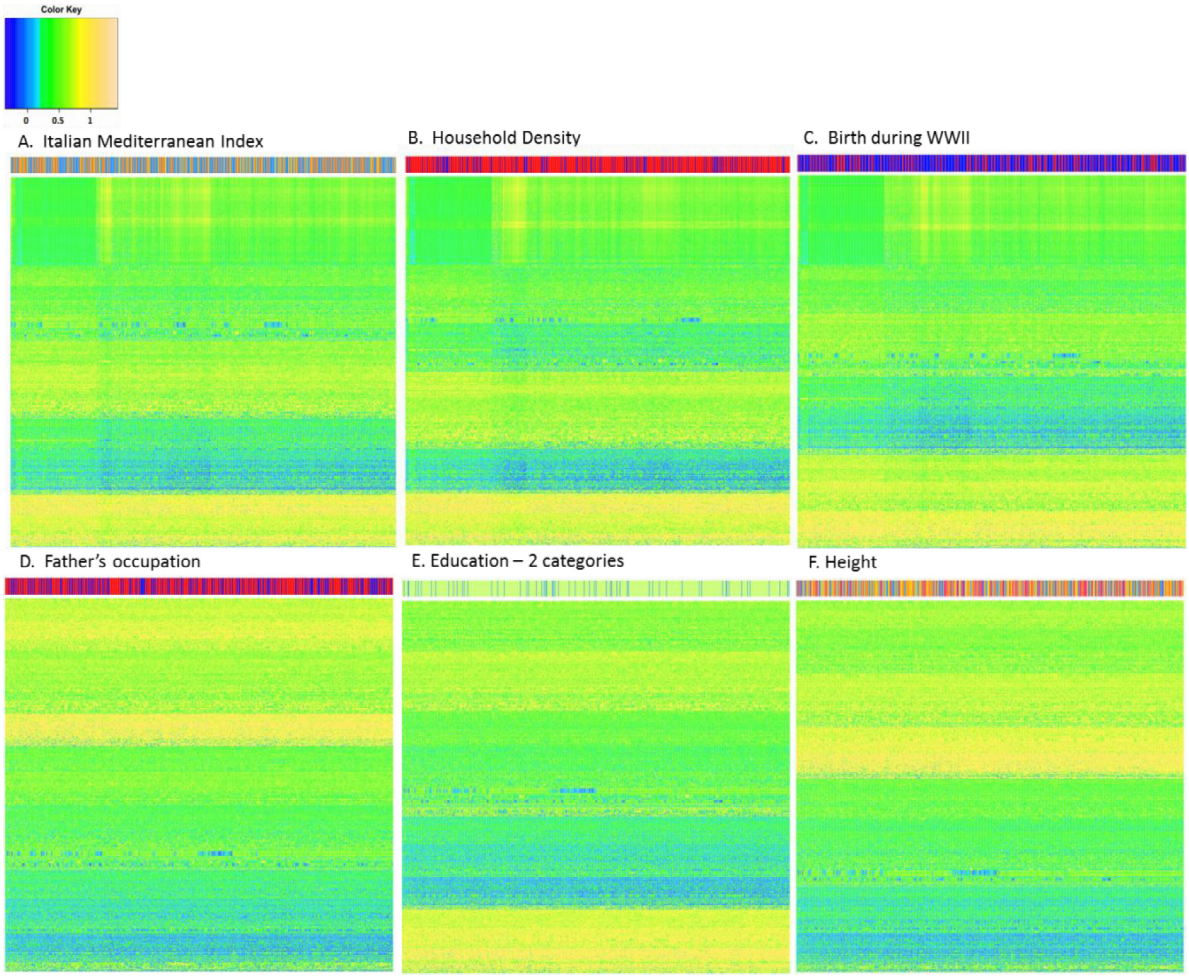


Figure 33 – Heatmaps displaying hierarchical clustering of subjects based on the 1000 most variable probes, using Euclidean distance with average linkage.

Samples are in columns (n=670) and CpG loci are in rows (n=1000). In the heatmap, blue indicates methylated and yellow indicates unmethylated CpGs. Failure to cluster subjects according to exposure is evident by the lack of clustering of colours on the column side bar where each colour indicates the exposure category to which the subject belongs.

Consensus clustering with a k-means algorithm and Manhattan as the distance metric was also used to cluster subjects, both for the whole as well as for the 'females only' dataset. For all subjects, male and female, the most robust number of clusters was identified as 2. Firstly, the heatmap produced with *pairwise consensus values* for

k=2 indicated a high stability of clustering since there was almost complete concordance between sub-sample clustering runs – subjects either always cluster together (deep blue colour) or are always not clustered together (white colour) (Figure 34A). Secondly, the empirical CDF plot indicated that for k=2 the CDF approaches the ideal step function with a step around 0 and then a second step at 1. The ideal CDF is that which is expected for perfect consensus where the histogram would have two bins, one centred at 0 (subjects that never cluster together) and the other at 1 (subjects that always cluster together) with an almost flat line in between (Figure 34B). Lastly, when k=2, cluster consensus for both clusters approached unity (Figure 34C).

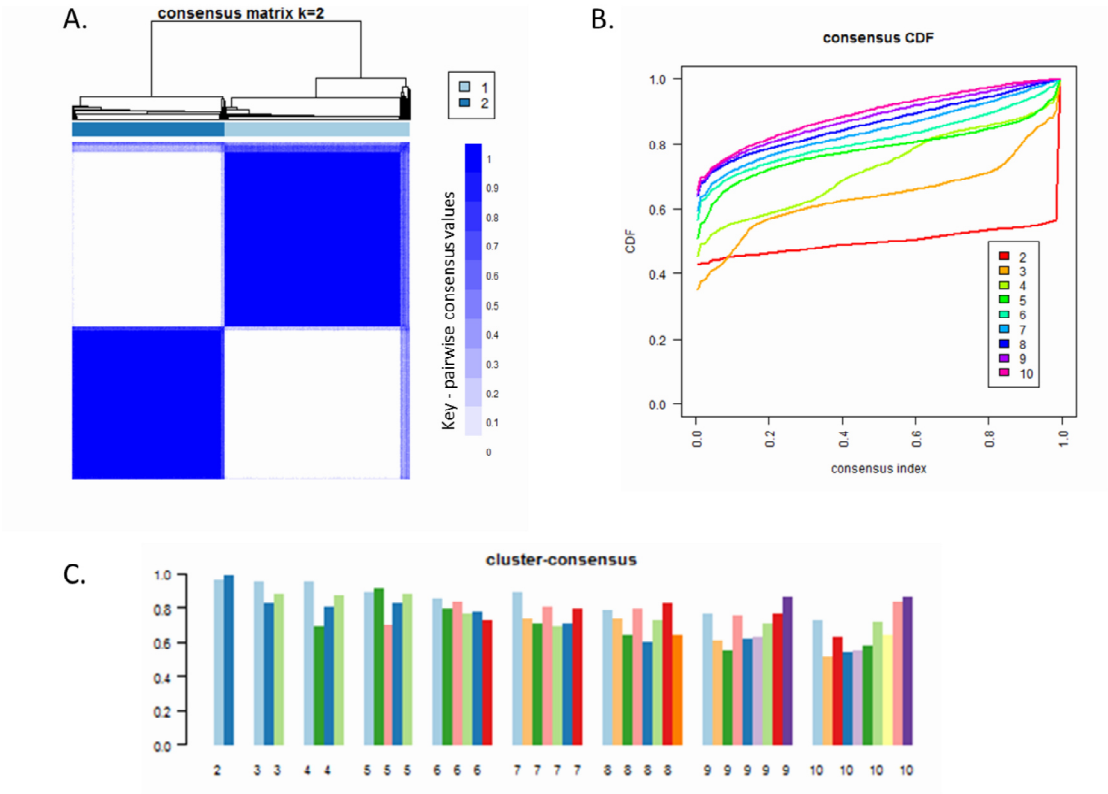


Figure 34 – Consensus clustering output graphics for the entire dataset (male and female subjects).

A. Consensus matrix heatmap indicating that when k=2 subjects either always cluster together between clustering runs (deep blue colour) or are never clustered together between clustering runs (white colour). B. Empirical Cumulative Distribution Function from a histogram with 100 bins displaying two steps for k=2, one at 0 and one at 1 with an almost flat line between. This is very similar to the CDF expected for perfect consensus. C. Cluster Consensus plot displaying the average pairwise consensus value of items in each cluster which, for k=2, is at unity for both clusters.

As evident in Figure 35, two was also the most robust number of clusters for the 'females only' dataset. The consensus matrix heatmap for k=2 had the cleanest cluster partition where there was almost complete concordance between sub-sample clustering runs. In addition, for k=2 the empirical CDF was closest to the ideal step function and when k=2, cluster consensus for both clusters approached unity.

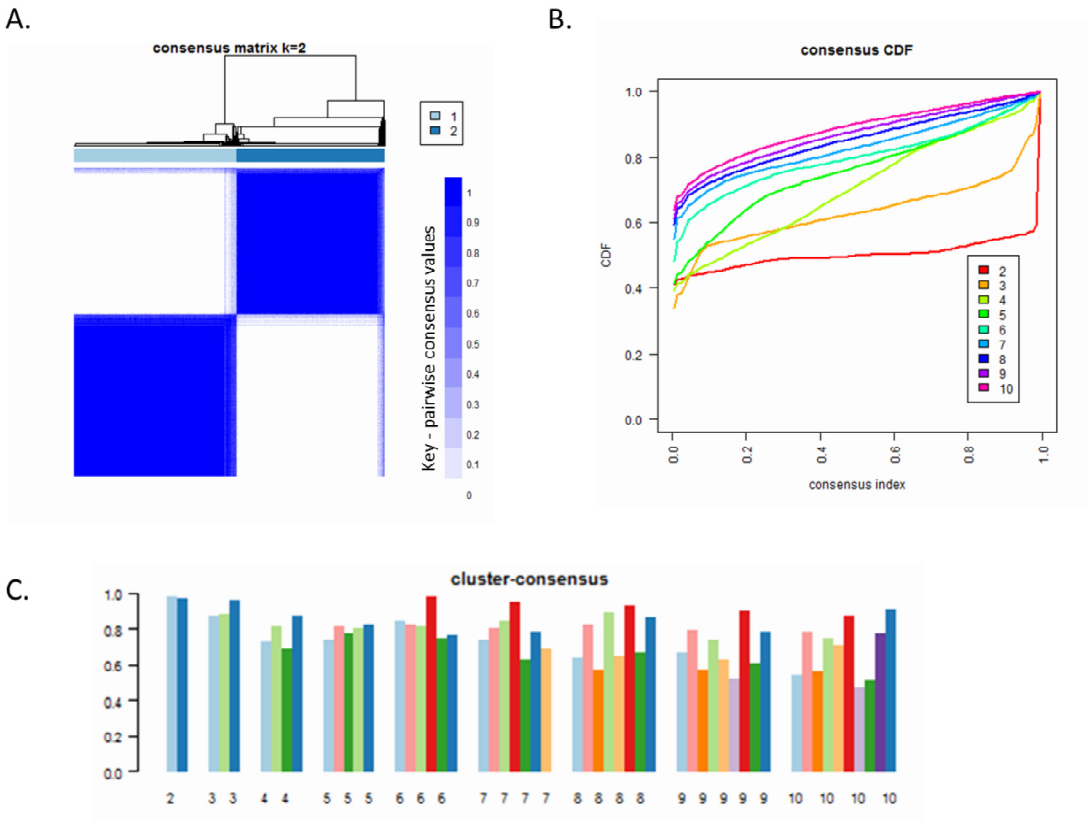


Figure 35 – Consensus clustering output graphics for the 'females only' dataset.

A. Consensus matrix heatmap indicating that when k=2 subjects either always cluster together between clustering runs (deep blue colour) or are never clustered together between clustering runs (white colour). B. Empirical Cumulative Distribution Function from a histogram with 100 bins displaying two large steps for k=2, one at 0 and one at 1 with a slightly upward line between. This is the closest empirical CDF to that expected for perfect consensus. C. Cluster Consensus plot displaying the average pairwise consensus value of items in each cluster which, for k=2, is at unity for both clusters.

However, similar to HC, the clusters derived from CC were not associated with any of the exposure variables examined, whether the exposure was relevant to all subjects or to female subjects only. Table 23 lists the P-value from a chi square test between cluster membership (whether the subject belonged to cluster 1 or 2) and exposure category, for each of the exposures of interest. Neither SES, nor anthropometry, nor nutrition, nor reproductive variables were shown to be significantly associated with cluster membership and thus overall methylation pattern.

Table 23 – Association between Consensus Clustering derived clusters and exposure category for each exposure variable of interest

Exposure variable	Association with CC cluster membership P-value[§]
Education (4 categories)	0.501
Education (2 categories)	0.451
Household Density	0.144
Household's highest occupational position	0.407
Father's occupational position	0.802
When household first had fridge	0.347
Birth during WWII	0.101
Italian Mediterranean Diet index	0.557
Total Caloric Intake	0.815
Alcohol Intake	0.996
Folate Intake	0.755
Methyl Intake Score	0.358
BMI	0.235
Height	0.478
Age at menarche[∞]	0.607
Menopausal Status[∞]	0.550
Age at menopause[∞]	0.119
Estimated Lifetime Oestrogen Exposure[∞]	0.192

[§] P-value derived from a chi square test between cluster membership (Consensus Clustering derived class 1 or 2) and exposure category, for each of the exposures of interest

[∞] The clusters used in this test were derived from consensus clustering of the 'females only' dataset

3.2.5 Principal Component Analysis

As illustrated in Table 24, the first PC from the PCA on the entire dataset was explaining 99.06% of the variance in our methylation data. This suggests that heterogeneity in the study population can be accurately summarized using a single axis. This first PC was found significantly associated with age, gender, dietary methyl intake, and caloric intake, in order of decreasing significance (P-values: <0.0001, 0.005, 0.046 and 0.050 respectively). For completeness, Table 24 also describes the two subsequent PCs (each explaining less than 0.1% of the variance). The second PC was significantly associated with gender, education, alcohol intake, height, and IMI score, in order of decreasing significance (P-values: <0.0001, 0.013, 0.019, 0.028 and 0.037 respectively). The third PC, was shown to be associated with age, gender, caloric intake, alcohol intake, height, year when family first had fridge and BMI (P-values: <0.0001, <0.0001, <0.0001, <0.0001, <0.0001, 0.002, and 0.009 respectively). Of the PCs illustrated in Table 24, only the third PC, which explains only 0.052% of the variance, evidenced an association with chip, reassuringly showing no significant technical variation (Table 24).

Similarly, the first PC derived from the PCA on the female subjects, explained 99.16% of the variance in the methylation data, whereas the second PC explained 0.068% and the third PC 0.019% of the variance. Same as above, none of these PCs was associated with chip. On the contrary, all three PCs were associated with age (P-values: 0.004, 0.008 and <0.0001 respectively) and menopausal status (0.024, 0.002, and 0.033 respectively). In addition, the third PC was also significantly associated with ELEE (P-value: 0.010) (Table 24).

Only the third PC from the PCA on the entirety of samples was found to be associated with chip. However, this PC explains only 0.052% of the variance in the dataset, and this shows that ComBat successfully corrected the vast majority of technical variance associated with batch effects.

Surprisingly, none of the PCs on either dataset was found to be associated with case-control status, suggesting that in these pre-diagnostic samples, age, gender and some of the risk factors for cancer, such as menopausal status, calories, and dietary methyl intake, might have more of an effect on methylation than future case-control status.

Table 24 – Associations between Principal Components and Subject Variables

PCA on entire dataset (n=667)			
	P-values from regression analysis		
	First PC	Second PC	Third PC
% of variance explained	99.06%	0.067%	0.052%
P-value for chip	0.466	0.124	<0.0001*
Variables:			
Age	<0.0001*	0.084	<0.0001*
Gender	0.005*	<0.0001*	<0.0001*
Case-Control status	0.461	0.924	0.546
Education (4 categories)	0.527	0.013*	0.503
Education (2 categories)	0.655	0.169	0.242
Household Density	0.267	0.484	0.548
Household's highest occupational position	0.283	0.193	0.070
Father's occupational position	0.665	0.691	0.822
When household first had fridge	0.757	0.307	0.002*
Birth during WWII	0.437	0.318	0.121
Italian Mediterranean Diet index	0.448	0.037*	0.210
Total Caloric Intake	0.050*	0.636	<0.0001*
Alcohol Intake	0.663	0.019*	<0.0001*
Folate Intake	0.883	0.392	0.418
Methyl Intake Score	0.046*	0.877	0.373
Height	0.763	0.028*	<0.0001*
BMI	0.095	0.643	0.009*
PCA on females-only dataset (n=485)			
	P-values from regression analysis		
	First PC	Second PC	Third PC
% of variance explained	99.16%	0.068%	0.019%
P-value for chip	0.160	0.504	0.608
Variables:			
Age	0.004*	0.008*	<0.0001*
Case-Control status	0.403	0.645	0.438
Age at menarche	0.548	0.209	0.077
Menopausal Status	0.024*	0.002*	0.033*
Age at menopause	0.887	0.057	0.563
Estimated Lifetime Oestrogen Exposure	0.106	0.128	0.010*

*P-value significant at the 0.05 threshold

3.2.6 Single Enrichment Analysis

Education (whether in 4 or 2 categories), household density, BMI, IMI, menopausal status, age at menopause, and ELEE were the exposures that produced negatively skewed P-value distributions suggesting a widespread effect on DNA methylation.

Of these exposures, only household density and ELEE produced at least 100 differentially methylated probes at the $1.0x10^{-5}$ threshold during locus by locus analyses and could thus be investigated with SEA.

Household density produced 152 differentially methylated probes with this relaxed cut off which corresponded to 110 unique genes. Functional classification of these genes produced only one cluster with an enrichment score ≥1.3, containing 8 genes relating to kinase structure and activity (Figure 36).



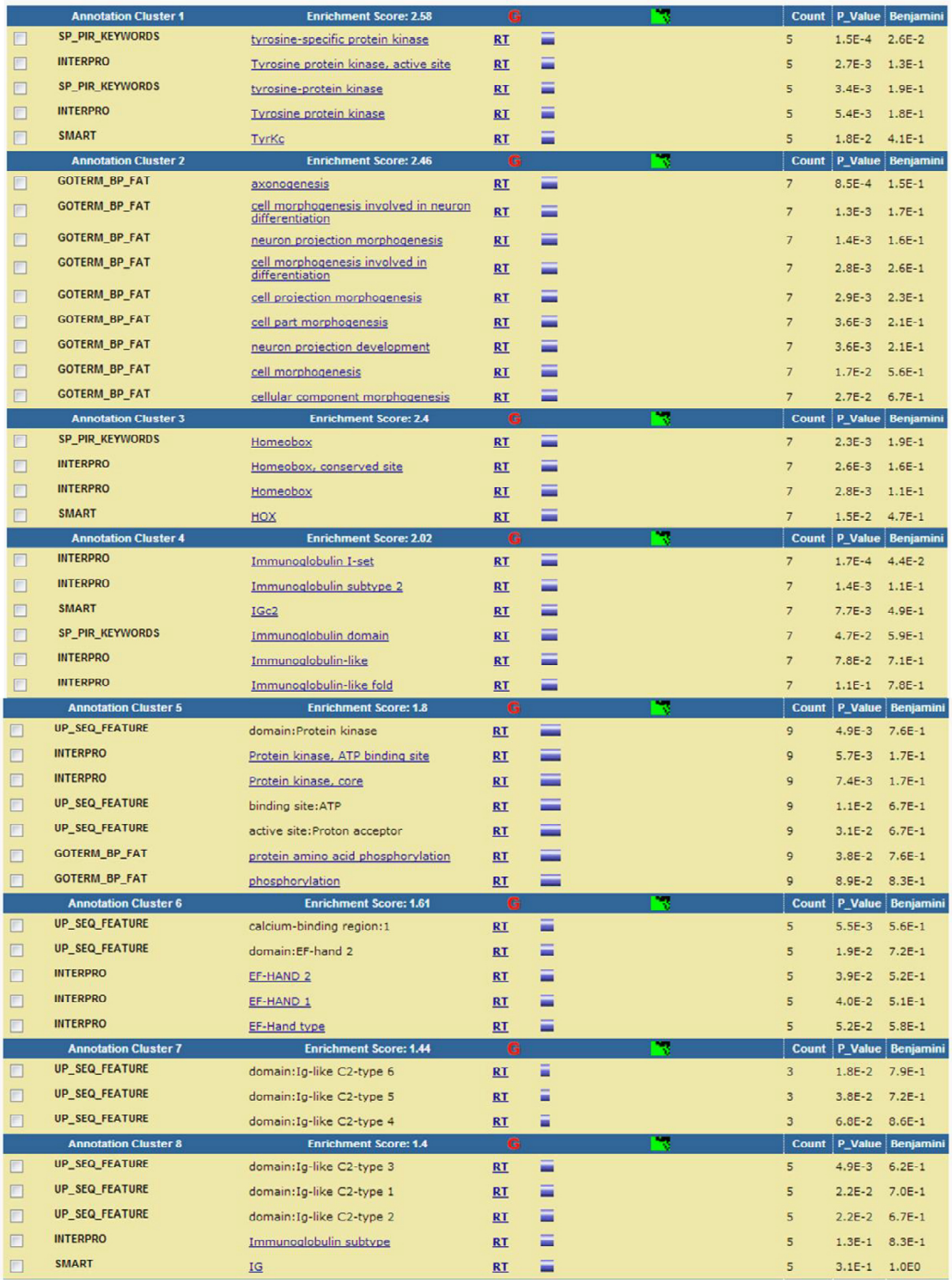
	Gene Group 1	Enrichment Score: 1.38 RG
1	MYLK3	myosin light chain kinase 3
2	IGF1R	insulin-like growth factor 1 receptor
3	ROR1	receptor tyrosine kinase-like orphan receptor 1
4	FLT4	fms-related tyrosine kinase 4
5	PRKD1	protein kinase D1
6	NUAK1	NUAK family, SNF1-like kinase, 1
7	KDR	kinase insert domain receptor (a type III receptor tyrosine kinase)
8	MST1R	macrophage stimulating 1 receptor (c-met-related tyrosine kinase)

Figure 36 – Functional classification of genes differentially methylated between household density exposure categories.

Highest stringency produced one cluster with enrichment score ≥1.3. Eight genes were included in this cluster all related to kinase structure and activity.

Functional annotation of the 110 genes revealed three biological terms that were significantly enriched at the Benjamini multiple-testing corrected P-value (≤0.05). These were a. transmembrane receptor protein tyrosine kinase activity, containing 6 genes – *CRIM1, FLT4, IGF1R, KDR, MST1* and *ROR*1 (Benjamini P-value: 0.013); b. tyrosine specific protein kinase , containing 5 genes - *FLT4, IGF1R, KDR, MST1* and *ROR1* (Benjamini P-value: 0.026); and c. Immunoglobin I-set, containing 7 genes – *FLT4, FSTL5, KDR, LSAMP, NFASC, PTPRS* and *ROR1* (Benjamini P-value: 0.044).

Lastly, functional annotation clustering produced 8 clusters with an enrichment score ≥1.3. The clusters are displayed in Figure 37. Cluster 1 encompasses biological terms that relate to tyrosine activity, cluster 2 to neuron development and cell morphogenesis, cluster 3 includes homeobox domains, cluster 4 relates to immunoglobulin structure, cluster 5 to protein kinase activity, cluster 6 to protein binding domains, and clusters 7 and 8 to Ig-like C2 domains. However, of all the biological annotations included in these 8 clusters only one, relating to tyrosine activity, had a significant P-value at the Benjamini corrected P-value cut off, casting doubt on the biological significance of the clustering results (Figure 37).



Annotation Cluster 1	Enrichment Score: 2.58			Count	P_Value	Benjamini
SP_PIR_KEYWORDS	tyrosine-specific protein kinase	RT		5	1.5E-4	2.6E-2
INTERPRO	Tyrosine protein kinase, active site	RT		5	2.7E-3	1.3E-1
SP_PIR_KEYWORDS	tyrosine-protein kinase	RT		5	3.4E-3	1.9E-1
INTERPRO	Tyrosine protein kinase	RT		5	5.4E-3	1.8E-1
SMART	TyrKc	RT		5	1.8E-2	4.1E-1
Annotation Cluster 2	**Enrichment Score: 2.46**			**Count**	**P_Value**	**Benjamini**
GOTERM_BP_FAT	axonogenesis	RT		7	8.5E-4	1.5E-1
GOTERM_BP_FAT	cell morphogenesis involved in neuron differentiation	RT		7	1.3E-3	1.7E-1
GOTERM_BP_FAT	neuron projection morphogenesis	RT		7	1.4E-3	1.6E-1
GOTERM_BP_FAT	cell morphogenesis involved in differentiation	RT		7	2.8E-3	2.6E-1
GOTERM_BP_FAT	cell projection morphogenesis	RT		7	2.9E-3	2.3E-1
GOTERM_BP_FAT	cell part morphogenesis	RT		7	3.6E-3	2.1E-1
GOTERM_BP_FAT	neuron projection development	RT		7	3.6E-3	2.1E-1
GOTERM_BP_FAT	cell morphogenesis	RT		7	1.7E-2	5.6E-1
GOTERM_BP_FAT	cellular component morphogenesis	RT		7	2.7E-2	6.7E-1
Annotation Cluster 3	**Enrichment Score: 2.4**			**Count**	**P_Value**	**Benjamini**
SP_PIR_KEYWORDS	Homeobox	RT		7	2.3E-3	1.9E-1
INTERPRO	Homeobox, conserved site	RT		7	2.6E-3	1.6E-1
INTERPRO	Homeobox	RT		7	2.8E-3	1.1E-1
SMART	HOX	RT		7	1.5E-2	4.7E-1
Annotation Cluster 4	**Enrichment Score: 2.02**			**Count**	**P_Value**	**Benjamini**
INTERPRO	Immunoglobulin I-set	RT		7	1.7E-4	4.4E-2
INTERPRO	Immunoglobulin subtype 2	RT		7	1.4E-3	1.1E-1
SMART	IGc2	RT		7	7.7E-3	4.9E-1
SP_PIR_KEYWORDS	Immunoglobulin domain	RT		7	4.7E-2	5.9E-1
INTERPRO	Immunoglobulin-like	RT		7	7.8E-2	7.1E-1
INTERPRO	Immunoglobulin-like fold	RT		7	1.1E-1	7.8E-1
Annotation Cluster 5	**Enrichment Score: 1.8**			**Count**	**P_Value**	**Benjamini**
UP_SEQ_FEATURE	domain:Protein kinase	RT		9	4.9E-3	7.6E-1
INTERPRO	Protein kinase, ATP binding site	RT		9	5.7E-3	1.7E-1
INTERPRO	Protein kinase, core	RT		9	7.4E-3	1.7E-1
UP_SEQ_FEATURE	binding site:ATP	RT		9	1.1E-2	6.7E-1
UP_SEQ_FEATURE	active site:Proton acceptor	RT		9	3.1E-2	6.7E-1
GOTERM_BP_FAT	protein amino acid phosphorylation	RT		9	3.8E-2	7.6E-1
GOTERM_BP_FAT	phosphorylation	RT		9	8.9E-2	8.3E-1
Annotation Cluster 6	**Enrichment Score: 1.61**			**Count**	**P_Value**	**Benjamini**
UP_SEQ_FEATURE	calcium-binding region:1	RT		5	5.5E-3	5.6E-1
UP_SEQ_FEATURE	domain:EF-hand 2	RT		5	1.9E-2	7.2E-1
INTERPRO	EF-HAND 2	RT		5	3.9E-2	5.2E-1
INTERPRO	EF-HAND 1	RT		5	4.0E-2	5.1E-1
INTERPRO	EF-Hand type	RT		5	5.2E-2	5.8E-1
Annotation Cluster 7	**Enrichment Score: 1.44**			**Count**	**P_Value**	**Benjamini**
UP_SEQ_FEATURE	domain:Ig-like C2-type 6	RT		3	1.8E-2	7.9E-1
UP_SEQ_FEATURE	domain:Ig-like C2-type 5	RT		3	3.8E-2	7.2E-1
UP_SEQ_FEATURE	domain:Ig-like C2-type 4	RT		3	6.8E-2	8.6E-1
Annotation Cluster 8	**Enrichment Score: 1.4**			**Count**	**P_Value**	**Benjamini**
UP_SEQ_FEATURE	domain:Ig-like C2-type 3	RT		5	4.9E-3	6.2E-1
UP_SEQ_FEATURE	domain:Ig-like C2-type 1	RT		5	2.2E-2	7.0E-1
UP_SEQ_FEATURE	domain:Ig-like C2-type 2	RT		5	2.2E-2	6.7E-1
INTERPRO	Immunoglobulin subtype	RT		5	1.3E-1	8.3E-1
SMART	IG	RT		5	3.1E-1	1.0E0

Figure 37 – Functional annotation clustering of genes differentially methylated between household density exposure categories.

Eight clusters with an enrichment score ≥1.3 were produced. However, the Benjamini P-value for individual biological annotations was only significant for one annotation.

Estimated lifetime estrogen exposure locus by locus analyses produced 115 differentially methylated probes with the $1x10^{-5}$ cut off which corresponded to 97 unique genes. Functional classification of these genes produced only one cluster containing 7 genes but the enrichment score of this cluster was 0.42 and was thus regarded as non-significant (Figure 38).

	Gene Group 1	Enrichment Score: 0.42	RG	T	
1	BEAN	brain expressed, associated with Nedd4			
2	FAM155A	family with sequence similarity 155, member A			
3	LYSMD3	LysM, putative peptidoglycan-binding, domain containing 3			
4	C4orf34	chromosome 4 open reading frame 34			
5	VOPP1	similar to EGFR-coamplified and overexpressed protein; EGFR-coamplified and overexpressed protein			
6	DISP2	dispatched homolog 2 (Drosophila)			
7	FNDC4	fibronectin type III domain containing 4			

Figure 38 – Functional classification of genes differentially methylated between estimated lifetime estrogen exposure categories.

Highest stringency produced one cluster containing 7 genes with an enrichment score 0.42, which was regarded as non-significant.

Similarly, functional annotation of the 97 genes differentially methylated between ELEE categories did not reveal any biological terms that were significantly enriched at the Benjamini multiple-testing corrected P-value (≤0.05).

Lastly, functional annotation clustering produced 4 clusters with an enrichment score ≥1.3 (Figure 39). Cluster 1 encompasses biological terms that relate to amino acid and fatty acid metabolism, cluster 2 to cell motion and migration, cluster 3 includes homeobox domains, and cluster 4 relates to hormone and endogenous stimuli response. However, similar to household density annotation clustering, none of the biological annotations included in these clusters had a significant P-value at the Benjamini corrected P-value cut off, casting doubt on the biological significance of the clustering results (Figure 39).

	Annotation Cluster 1	Enrichment Score: 1.87	G		Count	P_Value	Benjamini
	KEGG_PATHWAY	Tryptophan metabolism	RT		3	1.3E-2	6.7E-1
	KEGG_PATHWAY	Fatty acid metabolism	RT		3	1.3E-2	6.7E-1
	KEGG_PATHWAY	Lysine degradation	RT		3	1.5E-2	4.9E-1
	Annotation Cluster 2	**Enrichment Score: 1.82**	G		**Count**	**P_Value**	**Benjamini**
	GOTERM_BP_FAT	regulation of cell migration	RT		5	1.1E-2	9.5E-1
	GOTERM_BP_FAT	regulation of locomotion	RT		5	1.7E-2	9.0E-1
	GOTERM_BP_FAT	regulation of cell motion	RT		5	1.8E-2	8.7E-1
	Annotation Cluster 3	**Enrichment Score: 1.52**	G		**Count**	**P_Value**	**Benjamini**
	UP_SEQ_FEATURE	DNA-binding region:Homeobox	RT		5	1.4E-2	9.9E-1
	INTERPRO	Homeobox, conserved site	RT		5	3.1E-2	9.8E-1
	SP_PIR_KEYWORDS	Homeobox	RT		5	3.1E-2	9.3E-1
	INTERPRO	Homeobox	RT		5	3.2E-2	9.4E-1
	INTERPRO	Homeodomain-related	RT		5	3.3E-2	8.9E-1
	SMART	HOX	RT		5	5.1E-2	9.7E-1
	Annotation Cluster 4	**Enrichment Score: 1.48**	G		**Count**	**P_Value**	**Benjamini**
	GOTERM_BP_FAT	response to hormone stimulus	RT		7	1.2E-2	9.0E-1
	GOTERM_BP_FAT	response to endogenous stimulus	RT		7	1.8E-2	8.4E-1
	GOTERM_BP_FAT	response to organic substance	RT		7	1.7E-1	9.9E-1

Figure 39 – Functional annotation clustering of genes differentially methylated between estimated lifetime estrogen exposure categories.

4 Discussion

The table below summarises the results of this exploratory investigation by displaying which exposures have been shown to be significantly associated with methylation, and in which part(s) of the analysis (Table 25).

Table 25 – Summary of Results per exposure examined

	Method of Analysis					
Exposure variable	**LUMA global methylation**	**450k – Whole genome methylation**	**450k – Genome wide locus by locus analysis: top hits**	**450k – Genome wide locus by locus analysis: skewed P-value distributions**	**Consensus Clustering**	**PCA first Principal Component**
Education (4 categories)		√		√		
Education (2 categories)				√		
Household Density		√		√		
Household's highest occupational position		√				
Father's occupational position						
When household first had fridge						
Birth during WWII						
Italian Mediterranean Diet index				√		
Total Caloric Intake	√	√				√
Alcohol Intake						
Folate Intake	√	√				
Methyl Intake Score	√	√				√
BMI				√		
Height						
Age at menarche	√		√			
Menopausal Status						
Age at menopause				√		
Estimated Lifetime Oestrogen Exposure			√	√		

4.1 Discussion of Results

4.1.1 Socioeconomic Status

Early life SES, whether measured using father's occupation, when the family first had a fridge, or birth during WWII as proxies, was not associated with DNA methylation. Null findings were observed both using the LUMA assay and the Illumina Infinium 450 BeadChip, and also when the latter data were analysed as probe category specific methylation values, single loci, as clusters or as regressed against the PCA derived components.

Early life SES results contrast what was expected based on the available literature, where most studies investigating childhood SES, using various proxies, demonstrated significant effects on adult DNA methylation. As shown in Table 1, five studies evidenced a significant association between childhood SES and DNA methylation using, as proxies to childhood SES, early life SEP – a variable reflecting a combination of father's occupation and access to household amenities[312] – socioeconomic deprivation at birth [313], periconceptional exposure to famine [142,147], and early life poverty [314,318]. On the other hand, more in line with the results of the present analysis, are the results of investigations of global DNA methylation in response to prenatal nutrition and family SES, using surrogate measures of global methylation such as LINE-1 and Sat2 methylation, which did not yield any significant results [316–318].

The reason behind the null findings in this investigation could be the use of different assays and the lack of a valid proxy for the investigation of early life SES. For the LUMA investigation, only birth during WWII was available and could be used as a proxy to early life SES, but one cannot be certain that birth during the war truly did correlate with distress, and unfavourable socioeconomic circumstances. Furthermore, for the genome-wide methylation investigation, father's occupational position, manual or non-manual, time when household first had fridge, and birth during WWII were available as proxies to childhood social and economic conditions. It is possible, however, that none of these variables by itself accurately describes early life SES and instead, some combination of these variables, like in the investigation by Borghol et al. (2012) [312], would be more successful at characterising early life socioeconomic conditions.

In contrast to early life SES, in this investigation, adult life SES results were less consistent. Using LUMA, education was not found to be associated with global DNA methylation. Using the 450k BeadChip data, education, household density and highest household occupational position were found to be associated with methylation in different analyses.

The four different education categories were shown to have differential CpG island shore methylation, but this association was not significant at the Bonferroni corrected level. Similarly, education, whether examined in 4 or 2 categories, evidenced a heavily negatively skewed P-value distribution, even if no individual CpGs were identified as significant given the strict threshold of $<1x10^{-7}$. Education was also associated with the second PC, obtained from PCA on all subjects, but this PC only explained a negligible % of the methylation variance.

Very similar observations were made for household density which was shown to be associated with promoter and CpG Island methylation, but the associations did not retain significance at the Bonferroni corrected threshold. Household density also produced a heavily negatively skewed P-value distribution, even if no individual CpGs were identified as significant given the strict threshold of $<1x10^{-7}$.

Similarly, highest household occupational position was found to be associated with whole genome, promoter, CpG island, and CpG island shore methylation, only at the 0.05 level of significance. However, household's highest occupational position was not associated with differential methylation at any specific CpG, nor did it produce a skewed P-value distribution which would suggest a more diffuse effect on methylation.

These results might indicate that adult SES, measured through education, household density and/or highest household occupational position, affects DNA methylation patterns in a more diffuse way, but the limited power of this investigation – after correction for multiple testing – was not sufficient to detect individual differentially methylated CpGs. Any influence of adult SES variables on DNA methylation, however, was not sufficient to drive clustering, thus the clusters produced using the most variable probes were not associated with either exposure.

Single Enrichment Analysis on probes whose methylation seems to be influenced by household density, indicated that there was an over-representation of genes related to tyrosine protein kinase structure and activity. Functional annotation clustering also produced one tyrosine activity related cluster with significant annotation terms at the Benjamini corrected P-value. Also, SEA indicated that in these probes there was an enrichment of biological terms related to transmembrane receptor protein tyrosine kinase and tyrosine protein kinase activity. Tyrosine protein kinase is an enzyme that transfers phosphate groups from ATP to a tyrosine amino acid in a protein. This phosphorylation of proteins is important in protein activation and is essential in signal transduction[409]. Transmembrane receptor protein kinase is involved with transduction of extracellular signals to the interior of a cell, whereas tyrosine protein kinase is involved in signal transduction within the cell[409]. Signal transduction is essential in regulating several signal activities such as cell division, and inactivation of these proteins, for example due to epigenetic changes, may contribute to initiation or progression to cancer[409]. However, as stated earlier, a relaxed threshold of significance was used to create the gene list used in SEA and this list was fairly small, therefore these results need to be interpreted with much caution.

Adult SES was found to influence DNA methylation in 3 of the previous five studies in which it was investigated, all of which used different proxies for the exposure. Borghol et al. (2012)[312] estimated mid-adulthood socioeconomic position by combining the subjects' most recent occupation and household tenure. In this investigation, mid-adulthood SEP was associated with aberrant methylation of 545 promoters investigated. Another yet unpublished investigation, used household's highest occupational position as a proxy to adult SES and evidenced significantly influenced methylation at probes associated with pro-inflammatory genes, with low SEP predicting lower methylation [131]. Lastly, in a New-York based cohort, while income, education, and childhood SES did not predict repetitive element DNA methylation, an increase in wealth of 1 standard deviation was associated with very small and possibly biologically insignificant methylation changes (0.09% higher Alu and 0.15% lower LINE-1 methylation)

[318]. On the contrary, adult SES, as assessed using occupational prestige, was not associated with methylation differences in a community-based cohort [314] and education was not associated with global DNA methylation in peripheral blood in another investigation [313].

Overall, education was not shown to be significantly associated with DNA methylation at any of the available studies, all of which investigated global methylation measurements. This was in line with the LUMA results of this investigation. However, none of these studies investigated methylation at specific probe categories and hence they may have missed the association observed in the analysis of CpG island shore methylation and education. In addition, the available studies did not conduct a locus by locus analysis, which was found in this investigation to produce skewed P-value distributions for the education variables examined suggesting a more widespread effect on methylation.

In contrast to the study examining highest household occupational position, in this investigation, household's highest occupational position was only found to be associated with whole genome and probe type specific methylation, but not with specific CpG sites. The reason for this discrepancy is unknown but it might be due to the larger number of subjects in the unpublished investigation (n=857) and the candidate-gene approach which did not necessitate adjustments for multiple testing.

Lastly, none of the available studies investigated household density, therefore the results of this investigation, which seem to suggest an association between household density, as a proxy to adult SES, and DNA methylation, need to be strengthened, confirmed and validated in other investigations with increased power.

4.1.2 Nutrition

The investigation of dietary variables produced more consistent results. To begin with, using LUMA global methylation data, total caloric intake, folate intake and dietary methyl intake demonstrated significant associations. It was evidenced that highest caloric intake was associated with increased hypomethylation risk, whereas highest folate intake and increased dietary methyl intake score were associated with decreased hypomethylation risk. Simlarly, all three variables were shown to be associated with whole genome and/or probe type specific methylation, using the 450K BeadChip data. In addition, caloric, and dietary methyl intakes were associated with the first PC derived from PCA, suggesting that these dietary variables explain a large proportion of the variance in the methylation data. However, there were no specific CpGs whose methylation was significantly associated with these dietary variables, the P-value distributions did not suggest an effect on DNA methylation, and clustering of subjects using the most variable probes did not produce clusters associated with these exposures.

To the best of my knowledge, this is the first study that has investigated the effect of total caloric intake on DNA methylation. Therefore, the findings that increased caloric intake (>2215 vs. ≤1582 calories/day) is associated with increased risk of hypomethylation, using the LUMA data, and with whole genome as well as CpG island shore methylation, using the 450K data, need to be researched further using other assays in order to confirm whether these are true associations. It should be noted that the significant associations remained even after correcting for several related factors such as smoking status, BMI, total physical activity, folate intake and alcohol intake therefore it is unlikely that the results are due to confounding. Instead, it is possible that highest caloric intake, often associated with the consumption of energy dense foods and a high fat percentage, could have an effect on the epigenome, just like caloric restriction in utero leads to decreased placental DNA methylation [410], something that would suggest an inverse U-shaped curve between caloric intake and methylation. Nevertheless, the finding that highest caloric intake is associated with hypomethylation is supportive of a recent finding that, amongst premenopausal women, highest caloric intake significantly increases breast cancer risk [87].

The present findings on folate using LUMA, where highest folate intake (>335 vs. ≤225.4 μg/day) is associated with increased methylation, and the 450K BeadChip findings where folate intake is associated with whole genome and CpG island shore methylation, are supported by a number of studies which have demonstrated similar positive associations between folate intake or serum folate levels and global DNA methylation (Figure 14 and Table 2, Appendix D). In addition, the finding that a methyl intake score of above 0 – zero being below median intakes of folate and vitamins B1, B2, B6 and B12 in combination with above median consumption of alcohol – was associated with hypomethylation, and the finding that dietary methyl intake is associated with whole genome methylation, are also supported by relevant available literature, which explains the crucial roles of methyl nutrients in the proper maintenance of methylation in the cell [228,411].

Folate, alcohol and the B vitamins impact on one-carbon metabolism, the mechanism of methyl group transfers in the cell. Normal function of the one-carbon metabolism, central to maintaining a healthy epigenome, requires adequate supply of folate and the B vitamins. Folate depletion disrupts the delicate balance of the one-carbon metabolism and can sufficiently diminish levels of methylation in the cell [109,228,290,321]. In fact, different studies have reported diminished genomic DNA methylation in individuals with folate depletion [288–290]. In addition it was shown that folate therapy restores DNA methylation in individuals whose cells are hypomethylated [288–290,323].

On the contrary, alcohol has an adverse effect on the one-carbon metabolism, firstly by decreasing the bioavailability of folate. Heavy alcohol drinkers have been shown to have inadequate folate in their diet [337]. This is because alcohol causes malabsorption of folate due to loss of deconjugation – necessary activity at the colonic brush border – and partial loss of both folate binding proteins [338,339] and because alcohol causes increased renal excretion of folate due to a reduction in tubular reabsorption [340]. However, alcohol also impacts on the one-carbon metabolism by inactivating key enzymes in the pathway such as methyl synthase. This may lead to global hypomethylation or altered methylation in specific genes or loci [179].

Lastly, inadequate B vitamin intake has been shown to impact on global DNA methylation in animal models [412,413].

It is not surprising therefore that an association is observed between inadequate intakes of folate and dietary methyl intake and whole genome DNA methylation. The surprising finding is the lack of identification of differentially methylated CpGs in the 450k array, but the reasons for the discrepancy between the two assays can be due to the decreased power of the array to detect the associations after correcting for multiple testing. Also, it is surprising that clustering did not produce clusters relevant to these intakes, but it is possible that the methylation changes in response to these exposures are not sufficient to classify the affected probes in the 1000 most variable probes which were used for clustering.

With respect to alcohol, in all analyses carried out in this thesis, an association with DNA methylation, after correcting for folate intake, was not observed. This might contrast what would be expected according to the biological mechanisms described, but it agrees with the studies in Table 2, Appendix D. In the two studies that have investigated this association in blood, only the study investigating chronic alcoholism produced significant results [414]. These findings might reflect the possibility that alcohol exerts its effects through the influence of methyl group bioavailability and not directly on DNA methylation.

Lastly adherence to the Mediterranean diet was not shown to influence LUMA measured DNA global methylation. In addition, each subject's Mediterranean Diet Index score was not associated with whole genome or probe type specific DNA methylation, as measured with the 450K BeadChip array. Also, there were no probes whose methylation was significantly different between different degrees of adherence to the Mediterranean diet and clustering failed to cluster subjects according to Mediterranean diet adherence. However, IMI score did produce a negatively skewed P-value distribution which might indicate an effect on methylation which is insufficiently strong to be picked up by the power of this investigation.

The Mediterranean Diet is rich in fruits and vegetables which are themselves rich in folate, B vitamins and other methyl related nutrients. However, to delineate the true effects of the Mediterranean Diet and its components on DNA methylation, a more detailed and more powerful investigation needs to be carried out.

4.1.3 Anthropometry

Most of the results obtained on the anthropometric variables examined do not suggest an effect on DNA methylation. LUMA global methylation was not associated with height, BMI, waist-to-hip ratio or waist circumference. Similarly, genome-wide methylation did not evidence any associations with height or BMI. Also, neither variable was associated with whole genome or any probe category specific methylation measurements, nor were the variables associated with the clusters formed from clustering of subjects using the most variable probes, nor were the variables associated with the first PC which explains a large proportion of the variance in the methylation dataset. Similarly, no CpG sites were found to be significantly differentially methylated between anthropometric variable categories, and the P-value distribution for height is not different from what would be expected by chance. Only the P-value distribution of BMI is slightly skewed and seems to suggest a mild, widespread effect on methylation.

The results obtained are supported partly by the available studies investigating the association between BMI and other measures of adiposity and DNA methylation (Table 2) in peripheral blood. Of the 6 studies investigating BMI and peripheral blood DNA methylation in adults, only two studies evidenced statistically significant associations [362,363]. One study found a BMI of ≥25 compared to <25kg/m^2 to be associated with increased risk for below median LINE-1 hypomethylation in adult blood [363] while the other evidenced that for every unit increase in BMI, there was a 0.8% decrease in the methylation of the *MCHR1* gene. All other studies on adult BMI and blood DNA methylation evidenced null associations which are consistent with the results of this investigation. Most positive associations in the existing literature regard effects of overweight and obesity in infants and children which are considered periods of increased developmental plasticity, in tumours, and in response to weight loss interventions, all of which are outside the scope of this investigation.

Regarding waist circumference, there are only 2 available studies on peripheral blood methylation. One of these studies evidenced that a waist circumference >88cm vs. ≤88cm was associated with an increased risk for below median LINE-1 methylation [363] while the other, consistent with the LUMA findings in this investigation, did not produce a significant relationship between waist circumference and LINE-1 methylation[365].

To the best of my knowledge, this was the first investigation of the effects of waist to hip ratio on blood DNA methylation therefore there is no literature against which to compare the findings. Only one study investigating breast tumour tissue found that an increased waist to hip ratio was significantly associated with promoter methylation in at least one of the following genes: *E-cadherin, p16,* and *RAR-b2*[367].

With respect to height, there are no published studies investigating its effect on blood DNA methylation in adults. The only available studies are in children and adolescents, but their results might not be comparable to the results in adults. However, similar to the results of this investigation, two studies did not evidence a significant association between height and DNA methylation in adolescents and school-aged children [360,373]. Instead, of the three studies investigating height and DNA methylation in children (Table 4), the only significant association was

in a study which analysed cord blood samples and found methylation at specific loci to be predictive of height at 9 years of age[372]. Despite this intriguing finding, unfortunately this research question could not be examined in the available dataset.

4.1.4 Age at Menarche – Menopausal Status – Age at Menopause – ELEE

Reproductive variables also produced inconsistent associations with DNA methylation. Using the LUMA assay, only age at menarche was shown to be significantly associated with global DNA methylation. This association was not only strong but also demonstrated a dose-response relationship. However, using Infinium 450k BeadChip data, age at menarche was not found to be associated with DNA methylation – whole genome methylation or probe category specific methylation. Also, the overall P-value distribution produced from the locus-by-locus analysis of age at menarche and clustering of subjects using the most variable probes did not suggest a widespread effect on methylation. Instead, analyses identified 2 CpG loci differentially methylated between menarcheal age categories: cg27483455 and cg10903132. The first probe corresponds to a CpG site located in a CpG Island in the body of the *ACCN2* gene. This gene is a member of the acid-sensing ion channel (ASIC) family of proteins which have a function in neurotransmission and are involved with learning, pain transduction, touch sensation, and development of memory and fear [415]. This gene has not been found to be associated with age at menarche or cancer risk in genome-wide association studies [416,417]. The second locus, targets a CpG site in an intergenic region and its function is unknown. However, the body for this probe contains a sequence that matches to repetitive sequences in the genome[398], therefore the methylation value derived might not represent the actual methylation of the target CpG. None of these two CpG sites were previously found to be differentially methylated in different WBC types[401].

Despite the significance of these CpG loci, the differentially methylated CpGs demonstrated methylation differences between exposure groups that are well below the difference that can be detected with confidence, given the power of this analysis. In addition, the associations lose significance when examined only in women who remained healthy for the entire duration of follow up. These suggest that these hits could be false positive results or the result of confounding by case-control status. The lack of skewing in the age at menarche histogram might also indicate that the identified differentially methylated CpGs are false positive results. Further investigations, with increased power, are needed to determine whether these CpG sites are truly differentially methylated between exposure groups and whether such small methylation differences between groups are sufficient to have a biological effect.

The significant association observed using the LUMA data, agrees with the previous finding that age at menarche was associated with global methylation in a New-York City cohort [317]. However, the associations are not in the same direction. In the New-York cohort, a later age at menarche was associated with increased methylation whereas in this study it was observed that a later age at menarche was inversely associated with DNA methylation, as measured with LUMA. This association is also counter-intuitive since it contrasts the expectation that a protective exposure would be associated with increased methylation as breast cancer was shown to be associated with hypomethylation in various reports [348,418].

However, the negative association between LUMA methylation and later age at menarche, which decreases breast cancer risk, is consistent with the finding in the Long Island Breast Cancer Study Project (LIBCSP), where breast cancer was associated with increased global methylation as measured with LUMA [419]. This apparent paradox could be explained by LUMA's increased coverage in CpG dense regions, such as promoters, and decreased coverage in the remaining genome [379]. As previously discussed, breast cancer is characterized both by a global decrease in methylation as well as region specific, mostly promoter, hypermethylation. Therefore, promoter hypermethylation and LUMA's increased coverage in promoters may explain why there was an increased methylation detected in breast cancer cases compared to controls.

The conflicting findings between LUMA and Illumina 450k BeadChip methylation results in this analysis could be caused by differences in coverage between the two assays used, as explained earlier. LUMA assesses methylation of a specific restriction enzyme site (HpaII, CCGG), which occurs most frequently in CpG island promoters – also covered by the 450K array – and repetitive elements. However, the Infinium HumanMethylation 450 BeadChip, due to its probe design, does not interrogate repetitive elements. There are only a few probes that accidentally map to repetitive elements but these might not give an accurate picture of repetitive element methylation. If age at menarche disturbs methylation patterns in these repetitive elements, the hypomethylation would not have been evident in the 450K chip – which was not designed to cover repetitive elements – but it would have been detected in the LUMA assay. In addition, it is also possible that the lack of association in the 450K BeadChip was due to the decreased power of the array to detect the associations after correcting for multiple testing. Therefore, further investigations, with more power, into the role of oestrogen on methylation, and especially repetitive element methylation, are warranted.

The mechanistic link which might explain the association between menarcheal age and global DNA methylation is yet to be determined. However, endogenous oestrogen exposure is a strong candidate for epigenetic changes since an earlier age at menarche exposes a woman to a greater cumulative amount of endogenous oestrogens over her lifetime, and there have been reports which show that oestrogen impacts on DNA methylation. More specifically it was shown that oestrogen receptor (ER) positive breast tumour tissues have differential methylation at several CpG loci compared to ER negative tumours [420,421] and oestrogen induced breast tumours have differential DNA methylation patterns in ACI rat mammary gland tissue [422].

It is also possible that age at menarche is an indirect indicator of other macroscopic changes that may impact on DNA methylation. It is widely accepted that development is plastic and the sensitivity of the epigenetic system to environmental factors is heightened during periods of developmental plasticity such as childhood, adolescence and puberty [4]. Epigenetic modifications in response to environmental exposures at these critical periods are often subtle initially and even though they do not lead to phenotypic changes at the time of exposure, they may lead to increased risk of dysfunction and disease later on in life [4]. Trends in the past decades show a rapid shift towards an earlier age at menarche and this is more pronounced in developed countries [423]. This trend is too steep to be attributed to genetic changes. Instead, environmental exposures at the periods of developmental plasticity are likely to be the cause of the dramatic decrease in age at menarche. For example, childhood obesity disrupts the

hormonal milieu leading to an increase in adipocyte secreted leptin, or in adrenal secreted androgens, all of which impact on menarcheal onset [33]. Pre or neo-natal nutrition as well as early life exposure to endocrine disrupting chemicals (EDCs) can also influence age at menarche [4]. For example, a recent publication shows that early life dietary flavonol intake is associated with age of puberty in a cohort of girls[424]. Given that these exposures occur at the periods when the epigenetic signature is more plastic, they might also lead to aberrant DNA methylation changes which will be inherited during cell divisions and be detectable years later[147,358,403]. Therefore, the aberrant DNA methylation pattern observed in adulthood might not be related to menarcheal age per se but to an early life environmental exposure, such as early life dietary exposures, that impact both on age at menarche and on DNA methylation.

The LUMA results on age at menarche as well as a preliminary analysis of age at menarche on 332 female subjects using the HumanMethylation 450k BeadChip have recently been accepted for publication in PLoSONE journal (APPENDIX F). At the time the paper was written, only 332 female subjects were available for analysis and methylation data was analysed using PBC and Combat since BMIQ was not available. Nevertheless, using the 450K beadchip data, similar to this investigation, one CpG site on the *SMAD6* gene was identified as significant with a small percentage difference in methylation between exposure groups. This CpG site was not validated in an independent dataset using pyrosequencing, therefore, it was regarded as a false positive. Consequently, the conclusions of the manuscript are very similar to the conclusions presented in this thesis.

In view of all of the above, there is a need to carry out further studies with more power, better coverage and more detailed exposure assessment, to delineate the true relationship regarding the association between age at menarche and DNA methylation.

Regarding menopausal status, similar to the LUMA investigation, whole genome, probe category specific, and single locus methylation did not significantly differ between menopausal status categories. In addition, clustering of subjects using the most variable probes did not produce clusters associated with menopausal status. Only the P-value distribution demonstrated a slight peak at relatively small P-values but this slight skewness is not sufficient to suggest an overall effect of menopausal status on DNA methylation. However, menopausal status was found to be associated with all three PCs derived from PCA on the "females-only" dataset suggesting that menopausal status might explain a large percent of the variance. However, menopausal status is significantly correlated with age, and age was also associated with the PCA components. Therefore, it might be that age is producing the variance in methylation, and not menopausal status. So far, only one study has demonstrated an effect of menopausal status on the methylome but the sample size was small (n=162)[377]. Therefore, further investigations on this research question are warranted.

Menopausal age, above or below 50 years old, was not found to be associated with LUMA methylation measurements, or with whole genome, or probe category specific methylation measurements with the 450K BeadChip. In addition, menopausal age did not evidence any significantly differentially methylated CpGs. Only the P-value distribution from the locus by locus analyses was heavily negatively skewed suggesting that menopausal

age might leave a mark on the epigenome that is widespread but not significant to the point that can be detected with the power of this analysis. However, any effect of menopausal age on DNA methylation was not sufficient to drive clustering or explain a significant percent of the variance in this dataset. The mostly null results observed are in agreement with the only other study that has investigated age at menopause and found a non-significant association[418].

Similar to age at menopause, ELEE was not associated with global methylation using LUMA, or with whole genome, or probe category specific methylation measurements using the 450K BeadChip. ELEE did demonstrate a heavily negatively skewed P-value distribution in the locus by locus analysis suggesting that it might have a more widespread influence on DNA methylation. However, similar to menopausal age, the effect of ELEE on DNA methylation was not sufficient to drive clustering, or to explain sufficient variance in this dataset. ELEE was associated with the third PC from PCA on the "females-only" dataset, but this PC only explains 0.019%.

Despite the heavily skewed P-value distribution, only one CpG site was found to be differentially methylated between ELEE quartiles but the effect estimate was smaller than what can be detected with confidence in this analysis and probably too small to have a biologically significant effect. The differentially methylated CpG (cg01781745) is located in a CpG Island in the 1st exon of the *KLF9* gene. The encoded protein of the gene is a transcription factor whose binding activates or silences mRNA expression[425], however, no links between this gene and oestrogen exposure have been established. In addition, the association disappears when examined only in controls, pointing either to underlying confounding from case-control status or to limited power in the smaller subset to detect the association. This CpG site was not previously found to be differentially methylated in different WBC types[401].

When a relaxed significance threshold was used to run SEA on some of the probes that might be causing the skewed P-value distribution, in these probes there was no significant functional clustering of genes, no significant enrichment of biological terms and no significant functional annotation clustering. Therefore, no conclusions can be drawn regarding the function of the genes whose methylation appears to be influenced by ELEE.

Whether the significantly differentially methylated CpG and the skewed P-value distribution are true signals of biological variability or a false positive result arising by chance, needs to be investigated further in analyses with more power.

In these future investigations, a more accurate estimation of cumulative oestrogen exposure needs to be employed, in order to account for other factors influencing lifetime oestrogen exposure such as parity, lactation etc., since the definition employed here is only a crude estimation.

4.2 General Remarks

One of the most puzzling observations in this analysis was the lack of complete consensus between the LUMA global methylation and the 450k BeadChip whole genome methylation results. It was originally expected that the results would be in agreement, since both methods give an estimate of global methylation. However, it became evident that averaging the methylation of all probes did not give the same representation of global methylation as LUMA, since only the dietary variables produced consistent results between the two methods. As previously discussed, this might be because LUMA uses restriction enzymes that cut at CpG dense regions (islands in promoters and repetitive elements), with limited representation of the rest of the genome, whereas the 450K assay only has a few probes that match to repetitive elements of the genome and these may not give an accurate picture of repetitive element methylation. Hence the estimates of whole genome methylation derived are completely different (Median Methylation $_{LUMA}$: 71.8%; Median Methylation $_{450K\ whole\ genome}$: 0.574 or 57.4%) and it is no surprise that the same associations are not consistently observed.

Another striking observation was the reduced variance in the whole genome methylation measurements as measured with the HumanMethylation 450K BeadChip, compared to the LUMA global methylation measurements (Methylation SD_{LUMA}: 5.70%; Methylation $SD_{450K\ whole\ genome}$: 0.01 or 1.83%). This minimal variance suggests that whole genome methylation measured with the 450K BeadChip might not differ sufficiently between individuals, to enable us to observe significant variations between exposure groups. This reduced variance might also contribute to why some associations observed with LUMA, such as the association with age at menarche, are not replicated with the 450K BeadChip data.

Looking at the standard deviation of each probe, one can see that the median standard deviation is 0.018, which means that 50% of probes investigated have less than 2% variation in methylation. Assuming a power of 0.80 and an α-error of 0.05, in order to detect with confidence a 1SD (0.018) change in methylation of a particular probe between exposure groups, a very large sample size is needed (>10,000 subjects). In the investigations in this thesis, the differences in methylation between exposure groups are often just as small and are thus naturally not detected as significant given the power/sample size of this analysis. This confirms that, as previously discussed, the present investigation is of limited power to detect the subtle methylation differences that might exist between exposure groups. This is partly due to the limited sample size available but also due to the strict cut offs used when accounting for multiple testing.

The Bonferroni and FDR corrections used for this investigation are highly conservative corrections because of their assumption that all tests are independent. Literature suggests that this is not the case with methylation data since there is often correlation between methylation at neighbouring CpGs [426]. In order to get a more accurate cut off without having to make the 'independent tests' assumption, permutations[427] and the number of effective tests[428] have been suggested. However, in order to get an accurate threshold using permutation, at least 1000 permutations are needed for each analysis and given the number of probes and subjects as well as the number of

analyses carried out, this would be too computationally intensive and outside the time and technology constrains of this thesis. On the other hand, calculation of the effective number of tests requires calculation of eigen values though a PCA method which treats the probes as variables. However, PCA can only give accurate estimates when the number of variables is less than the number of subjects and in the present investigation, there are close to half a million probes but only 667 subjects. This would mean that the maximum number of effective tests that could be identified is 667, but even these 667 tests only explained a very small proportion of the variance in the data. As a result, this method cannot provide an accurate number of the effective tests and hence it cannot provide a less strict multiple testing correction.

Regarding the P-value distributions obtained, there is some suggestion that variables such as education, household density, IMI, BMI, menopausal age, and ELEE have a more widespread effect on methylation that is not detected as significant using the strict cut-offs employed for multiple testing correction. This suggestion arises from the difference of the P-value distributions obtained from those that would be expected by chance under the assumption that each test (and therefore each probe) is independent. However, as previously discussed, each test is not independent because of correlation between methylation at neighbouring CpG sites. Permutation would help establish P-value distributions that truly reflect the distribution of P-values if the null hypothesis for each exposure was true. However, due to the limited time constraints that do not allow permutations to be run, one cannot be certain what the null P-value distributions really look like, which diminishes the importance that can be placed on these results.

The null results obtained in clustering analyses could be explained by the small standard deviations in methylation for each probe in the 450k BeadChip data. This minimal variance means that even the most variable probes are not markedly differently methylated between subjects and, consequently, between exposure groups. Therefore it is no surprise that clustering using these probes failed to cluster individuals according to the exposures examined. Even the exposures that were suggested to leave a mark on the epigenome, did not impact on probe methylation sufficiently to produce large enough differences to drive clustering in relation to exposure groups.

In addition to the limited power of the investigation and the limited variance in methylation for each probe, another explanation for the mostly null results obtained in the Infinium 450K analysis, is the fact that this thesis investigated exposures that are likely to have subtle effects at the molecular level and these effects might even be less than the biological variability in the tissue examined, in this case blood. On the contrary, exposures with stronger molecular effects, such as smoking, were shown to be associated with blood DNA methylation in other investigations [291]. Also, as it became evident from the PCA results, gender and age explain a large percentage of the variance in our datasets. PCA results suggest that almost all variance in the 450K datasets is explained by age and gender, and the strong effects of these variables might obscure the more subtle effects of the exposures examined here.

Lastly, the results of SEA need to be very cautiously interpreted as previously mentioned, because their accuracy and reliability are questionable. In order for SEA to produce reliable results, the gene list used needs to contain

many important genes (marker genes), a large number of genes (>100) and most of the genes need to be differentially methylated with respect to the exposure or outcome examined and their differential methylation needs to be well above the threshold of significance [406]. In our investigation, almost no genes were found to be differentially methylated above the threshold of significance and the genes entered into SEA were selected following a relaxing of the significance threshold. Still, even with this relaxed threshold, only two exposures – household density and ELEE – had more than 100 differentially methylated probes. Also, most of the clusters produced in functional annotation clustering for both exposures lacked significance at the Benjamini corrected threshold of significance. As a result, the SEA results need to be treated with extreme caution as they might not be a true representation of the biological significance of methylation changes in the cell.

Despite the limited power and the lack of significant results, especially with the 450k BeadChip array, this was a preliminary analysis which gives us some indications that can shape and direct future investigations. One can regard this investigation as a first descriptive approach, which points to those exposures that need to be researched in more detail and with more power in the future.

Even though an attempt has been made above to compare the results of this investigation to other published studies, great care needs to go into the interpretation of these comparisons because the assays used in this investigation have not been used by other studies included in this thesis, with the exception of one [316]. Instead, global methylation was usually measured using pyrosequencing of repetitive elements (LINE-1 and Alu) whereas locus-specific methylation was usually measured using other arrays or methylation-specific PCR or bisulfite sequencing. Since different assays use different technologies, they might give different representations of methylation and with different technical variability making direct comparison of results difficult.

4.3 Difficulties in Establishing Causality in Environment-Epigenome Investigations

Even though environmental exposures have been shown to impact on DNA methylation, evidence to support causal relationships between exposures and epigenetics is lacking. There are several reasons why this is the case. Firstly, any exposure examined is only a small fraction of the totality of all exposures that an individual is exposed to from conception to death – defined as the exposome [429]. The exposome includes internal exposures such as metabolism, hormones, and inflammation, markers of general external exposures such as education and financial status, as well as specific external exposures such as infectious agents, lifestyle factors, and diet [429] (Figure 40). It is also true that these three domains of the exposome are not mutually exclusive. For example, a high fat diet during childhood which is considered a specific external exposure can upset [59] the hormonal balance which is an internal exposure. In addition, education and socioeconomic status can influence one's dietary choices. However, the interplay between exposures is almost impossible to investigate accurately, making inferences about causality difficult.

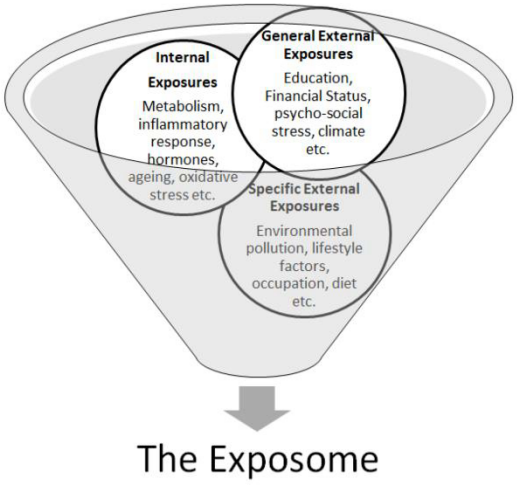


Figure 40 – The complementarity of the exposome

Secondly, when studying exposures, one should not only consider the interplay between the three domains of the exposome and the countless components of each, but also the dynamic nature of the exposome. The timing, duration, and intensity of the exposure are of extreme importance in these investigations since some phenotypic changes may not be evident after short term or weak exposures. Not accounting for these factors in investigations may obscure any true effects. One key issue to consider here is that the susceptibility of the epigenome to internal as well as external influences may be greater in critical windows during development. For example, during embryonic development DNA methylation marks are almost completely erased and then re-established [430]. Any environmental influences, such as carcinogen exposure or inadequate intake of methyl groups, are likely to have a substantial impact during this period of dramatic epigenetic reprogramming, and any effects can persist and alter disease risk later on in life [167]. As a result, in order to get an accurate picture of the interplay between an exposure and the epigenome, precise assessment of the exposure at different time-

frameworks to determine when the exposure is more aetiologically relevant is warranted. Furthermore, different metrics of exposure such as duration of exposure in addition to degree of exposure need to be investigated.

A further complexity in environment-epigenome investigations is that the epigenome, including DNA methylation, is dynamic with often dramatic cycles of demethylation and re-methylation. This plasticity and reversibility of the epigenome makes identification of adverse changes associated with particular exposures difficult. For example, many studies investigate changes in protein levels and in the activity of epigenetic players in response to external stimuli. However, one cannot be certain that changes in these players will have adverse phenotypic consequences on the epigenome. This might be because these changes are transient in nature and will not have a true lasting effect on the epigenome state, or they may be part of the normal adaptive response to stimuli [167].

Another issue that complicates investigations of causative associations between exposures and the epigenetic state is that an exposure may affect several of the different epigenetic layers in different ways, and inevitably, depending on the epigenetic outcome of a particular study, other effects might be missed. In addition, despite the progress in methylation interrogating techniques available, most cannot differentiate between 5mC and 5hmC bases and for this, no studies or techniques have to date addressed methylation changes at the 5hmC level [167]. Just like environmental exposures can impact on DNA methylation, 5hmC effects are also possible.

It is also important to consider that any changes observed in the epigenetic state following exposure to external or internal stimuli might not be caused directly from the exposure but they might be correlated with the adverse phenotype produced by the exposure [167]. Therefore, it is important to differentiate the driving and functionally important epigenetic changes from the passenger epigenetic consequences of adverse phenotypes such as malignancy [183,311]. More specifically, in retrospective studies one should consider whether the changes observed are a result of treatment rather than disease, especially since many retrospective studies recruited post-treatment subjects/samples. Similarly, in prospective studies one needs to consider the long latent period of various diseases and especially cancer, since undiagnosed disease might have an impact on DNA methylation which might be wrongly interpreted as response to external exposures.

Furthermore, as demonstrated by Brennan and Flanagan (2012) [431] the choice of technique for measuring methylation, and especially global methylation, might result in non-representative estimates of true DNA methylation change, obscuring any true effects. For example, even though 5mC content reveals a global hypomethylation in blood of cancer cases compared to controls, LINE-1, as a surrogate for global methylation, fails to demonstrate this association[431].

Another important issue that needs to be accounted for is technical variability. Statistical analyses with sufficiently large sample sizes, may produce statistically significant findings in which the effect estimates are smaller than the technical variability of the assay used to derive the methylation measurements [431]. However, significant findings with small effect estimates may not be reproducible using other technologies (for example pyrosequencing, one of the most quantitative assays available, can only detect changes greater than 3%) [431].

Additionally, tissue heterogeneity is one factor that can confound investigations on DNA methylation, making the establishment of causal associations difficult. It has been shown that DNA methylation differs between different tissues [146] and even different white blood cell types have variable DNA methylation patterns [293,314,400,432]. Thus, when the sample under investigation, whether tumour/normal tissue or blood, is a heterogeneous mixture of various cell type populations, variability in cell type counts between samples can account for a considerable proportion of variability in DNA methylation. Therefore, accounting for tissue heterogeneity in DNA methylation investigations where the exposure or outcome could alter cell type counts is imperative.

Lastly, perhaps the greatest limitation which precludes causal inferences in such investigations is that what constitutes a "normal" epigenome has not been established yet and thus it is difficult to distinguish pathogenic changes in response to the exposure from normal variability or even from adaptive response. Without a notion of what is "normal" it is also impossible to account for individual susceptibility which is highly variable and may mask any significant effects when subjects are pooled together [167]. International initiatives powered by the genome/epigenome technology revolution will hopefully shed light on what constitutes a "normal" epigenome and what is regarded as "normal" variability [167].

These and other issues arising in the investigation of causal exposure-epigenome investigations are discussed in a joint publication between members of my department at Imperial College and researchers at the International Agency for Research on Cancer. The manuscript which has been accepted and is awaiting publication in Carcinogenesis is attached in Appendix E.

4.4 Limitations of the Investigation

In addition to the caveats reported above, this investigation also had some additional limitations which could have precluded us from detecting any real associations. Firstly, blood is a surrogate tissue for many diseases such as cancer. The exposures examined in this thesis have been shown to influence breast cancer risk and if they do so through changes in the methylome, then such methylome aberrations would be mostly evident in the target tissue, breast. Therefore, even if these exposures did impact on disease risk though epigenetic changes, one might not see these changes in surrogate tissues, such as blood.

Also, as mentioned above, blood is a highly heterogeneous tissue and different blood cell type counts between samples could dilute estimates of differential methylation between exposure categories or produce false positive estimates. Unfortunately at the time of conduction of this investigation, even though an algorithm which uses DNA methylation signatures to infer cell mixture distribution[433] was available, purified cell samples to be used for the development of the validation set were not available. The only available option to account for this problem was to exclude from our analyses those probes that were previously shown to be differentially methylated between white blood cell types [401]. However, it was comforting to see that the results with and without the white blood cell type differentially methylated probes did not differ. This was not surprising as many of the exposures analysed, such as age at menarche, anthropometry and nutrition are not very likely to influence blood cell counts.

Furthermore, methylation changes in response to external exposures are probably cell specific and dose and time dependent, since there may be different mechanisms in different tissues conferring different susceptibilities to different exposures [434]. Since blood cells are mostly quiescent and post-mitotic, methylation levels might be more stable in blood than in other actively dividing cells [434]. Therefore, one might question the appropriateness of blood for monitoring variations in methylation in response to these external exposures. In fact, the use of blood DNA could be one explanation for the very limited variability in DNA methylation measurements between subjects and exposure groups. In a publication examining tissue specific DNA methylation changes in response to age, even though 274 CpG sites were shown to be differentially methylated with respect to age in solid tissues, in blood, there were zero differentially methylated CpGs[146].

Lastly, the assays used are not without limitations. The LUMA assay, due to the assessment of methylation at a specific restriction enzyme site (HpaII, CCGG), which occurs most frequently in CpG island promoters and repetitive elements, might not give an accurate estimate of global methylation. It is not surprising that while most studies on blood report global hypomethylation in response to breast cancer[171,431], in a large case-control study where methylation was assessed with LUMA, breast cancer was associated with increased genome wide methylation[419].

There are also limitations associated with the HumanMethylation450 BeadChip. Despite going into great lengths to correct for the differential probe design, having two different types of probes introduces variability into the methylation measurements which could cloud up any biological differences between exposure groups. On the

other hand, one cannot ignore the possibility that normalisation techniques might lead to over-correcting, thus diluting any true biological signals. In addition, as discussed in the methods section, the design of a large number of the probes on the array is imperfect. Expanded biological annotation of the array revealed that 4.3% of the probes had a CpG target site which was a documented SNP and 8.6% of probes mapped to multiple locations of the genome as assessed *in silico* [398]. Excluding these probes, in addition to the probes which failed in more than 20% of samples and the probes containing a SNP in their body, resulted in a substantial reduction in coverage (around 300, 000 probes). On the other hand, retaining these probes could have confounded the analyses.

4.5 Future Directions

The difficulties in establishing causality outlined above, as well as the limitations of this investigation, point to the need for better, more detailed and more suitable study designs in order to understand the effect of these and other exposures on DNA methylation.

In order to gain a better understanding of methylation effects given the dynamic nature of the exposome and the importance of timing, duration, and intensity of exposures, longitudinal and prospective cohort studies need to be employed[403]. These study designs, accompanied by serial sampling, allow assessment of changes in exposure and in DNA methylation throughout the life-course [435]. This in turn will allow identification of the critical windows for each exposure. In other words, these studies allow one to define at which time-points each exposure has most of an effect on DNA methylation. Prospective studies will also allow investigation of the relationship between aberrant methylation due to exposures and disease risk[403,436].

As previously discussed, early development is considered a critical window of exposure with respect to DNA methylation, because the plasticity of the epigenome is heightened during this early life period. As a result, birth cohort studies will be invaluable in researching the effects of early life exposures on DNA methylation, both during development but also in adult life. Birth cohort studies, just like all cohort studies, will also enable an evaluation of the DNA methylation changes on adult disease risk.

Perhaps the most informative study design for the hypotheses under question is a longitudinal analysis of epigenetic changes in a population cohort of monozygotic twins, with serial sampling [403]. Monozygotic twins share the same DNA sequence and thus any substantial disparate epigenetic changes point to non-shared environmental exposures rather than genetic influence[403,437]. Assessment of exposures and DNA methylation pattern at different time-frames throughout life can provide valuable clues as to which exposures could have precipitated an aberrant DNA methylation pattern.

If budget and time constrains do not permit such studies, monozygotic twins, discordant for the exposure of interest can also be used, since any epigenetic alterations would be environmentally rather than genetically driven[403]. However, even if the twin pairs are not recruited at birth or during early life, information on early life exposures would be essential in order to correct for confounding and investigate the true effects of the exposure under question.

Despite the different study designs available, in epigenetic epidemiology it is difficult to select a gold-standard study design. Instead, it is accepted that "combined analysis of multiple studies with complementary strengths will be necessary to identify meaningful and interpretable epigenetic associations"[403].

In addition to the study design, careful consideration of the analytical methods to be employed is imperative for the successful investigation of the effect of exposures on DNA methylation.

Firstly, despite its high cost, bisulphite conversion followed by whole genome sequencing permits investigation of methylation at all ~28million CpG sites, which is a vast improvement from the 450k BeadChip array which interrogates <2% of the CpG sites in the human genome [403]. Therefore, whole genome sequencing will play a major role in future epigenetic investigations, especially as next generation sequencing technologies become more available and more cost effective[436].

As already commented above, there are numerous combinations of pre-processing methods to be used on high throughput data. In addition, given the cost of high throughput technologies, the sample sizes analysed are often insufficient to detect the subtle effects that these exposures are likely to have on the epigenome. Therefore, creation of consortia, with pooling of subjects analysed with the same technology platform, and consensus agreement on the most appropriate combination of pre-processing methods to be used, would enable detection of differentially methylated sites or regions with respect to exposures, even if the exposure effects on methylation are subtle [403].

Another analytical consideration that might be useful in future epigenome-environment interactions is the analysis of genes based on functional annotations[403]. The ability to combine CpG sites, genes, or regions according to their function enables fewer and more biologically meaningful comparisons, which minimises multiple testing issues. However, this analysis is a sort of candidate approach and requires better understanding of the epigenome-environment interactions and improved identification of the locations of exposure driven variation[403,438]. A recent study revealed that only 21.8% of CpGs show developmental variation and these CpGs were shown to co-localize with gene regulatory elements, providing a starting point for more effective and focused investigations into environmental effects that could lead to developmental alterations and carcinogenesis [408].

Future investigations also require an association of DNA methylation with transcriptional activity [403]. Just as in this investigation, the methylation changes observed with respect to exposures are often subtle and it is not clear whether the magnitude of the change is enough to have an effect on transcription. Only if an association between methylation change and change in transcription is established, can the methylation change be regarded as of functional relevance [403]. Therefore, future studies should seek to establish such associations that would confirm functional relevance, bearing in mind the tissue specificity of expression.

In addition, Mendelian randomisation (Figure 41) can be an invaluable epidemiological tool in future studies to strengthen any conclusions regarding causal inference[403,439]. The first step of epigenetic Mendelian randomisation uses a genetic variant as a proxy to the environmental exposure and assesses the relationship between this genetic variant and the methylation outcome [439]. A number of validated genetic variant proxies for exposures such as smoking, alcohol and BMI exist and can be utilised in these investigations, provided that they do not have a direct influence on DNA methylation [439]. For example, maternal *MTHFR* genotype was suggested as a proxy to reduced folate availability in utero[440]. Investigation of the genetic variant instead of the environmental exposure deals with the common issues of confounding, reverse causation and measurement error, and thus produces

more accurate effect estimates. If there is agreement between the effect estimate of the association between the genetic variant used as proxy to the exposure and DNA methylation with the estimate of the association between the exposure and the DNA methylation, then the observed association can be regarded as an unbiased, unconfounded causal effect [439]. However, Mendelian randomisation can take these epigenome-environment interactions a step further, since in step 2, a genetic variant which can be used as a proxy to DNA methylation, is assessed for its association with a disease of interest[439]. A lot of work needs to go into identifying suitable genetic variants for different DNA methylation outcomes, but this two-step epigenetic Mendelian randomisation can provide evidence regarding whether an environmental exposure is causally associated with disease and whether the underlying biological mechanisms involve the epigenome.

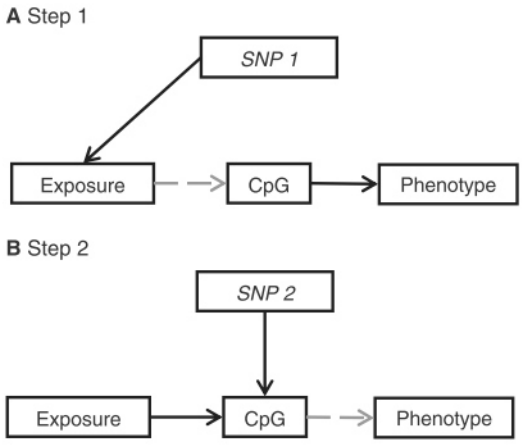


Figure 41 – Two-step epigenetic Mendelian randomisation: applying the principle of Mendelian randomisation to DNA methylation as an intermediate phenotype.

Genetic variants can be used as instrumental variables in a two-step framework to establish whether DNA methylation is on the causal pathway between exposure and disease. An overview of the two-step framework of this approach is shown. (A) First, a SNP is used to proxy for the environmentally modifiable exposure of interest and (B) secondly, a different SNP is used to proxy for DNA methylation levels. Reproduced with permission from [439].

Lastly, approaches that combine study designs can also support causal inference in the investigations between exposures, DNA methylation and disease. Vineis and Perera (2007) [441] propose the use of complementary study designs to establish new biomarkers, such as DNA methylation changes, that lie on the causal pathway between exposure and disease. For example, methylation markers that in prospective studies are shown to be associated with the exposure of interest and are also shown to be predictive of disease risk, can be classified as biomarkers of the mechanistic pathway linking exposure to disease, and the association between exposure and disease can be regarded as causal (Figure 42). This approach, termed “Meet-in-the-Middle”, can open up new avenues for establishing causal relationships which will then shed light on possible targets for prevention and treatment. Thus, “Meet-in-the-Middle” will be invaluable in future research efforts for establishing epigenome-environment interactions and their implications in disease risk.

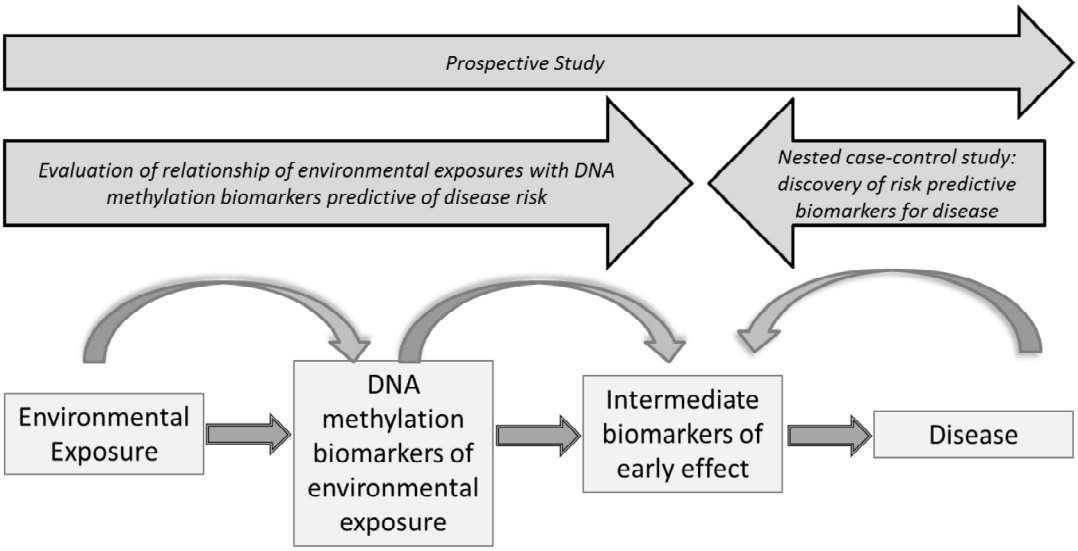


Figure 42 – The Meet-in-the-Middle approach: an epidemiological tool to reinforce causal inference in epigenetic epidemiology.

The figure demonstrates the use of complementary study designs to identify DNA methylation biomarkers that lie on the causal pathway between exposure and disease.

Combinations of the study designs and analytical approaches above produce great anticipation for the future of epigenome-environment interactions, as they provide the means through which the full potential of epigenetics – in prediction, prevention, and treatment [435] – can be uncovered.

Therefore, a re-analysis of the exposures examined, using combinations of these study designs and analytic methods, would be needed to confirm the associations observed in this analysis and also shed more light on the exposures which currently produce inconclusive results.

5 Conclusions

In this exploratory investigation of environmental exposures undergoing secular trends, exposures such as age at menarche, folate intake, methyl group intake, and total caloric intake were shown to influence global DNA methylation using the LUMA assay. In addition, education, household density, highest household occupational position, total caloric intake, folate intake, and dietary methyl intake were also shown to influence whole genome and/or probe category specific methylation using the HumanMethylation 450k BeadChip. Total caloric intake, dietary methyl intake and menopausal status were also found to, along with age and gender, weigh on the first PCs that explain more than 99% of the variance in the 450K datasets. Lastly, with the same array, there are suggestions that exposures such as education, household density, adherence to the Italian Mediterranean diet, BMI, age at menopause and lifetime oestrogen exposure affect the methylation of specific CpG loci but the power of this investigation was not sufficient to identify specific biologically relevant, differentially methylated CpGs.

Total caloric intake, folate intake, and dietary methyl intake are exposures which were found to be associated with methylation using both assays in this thesis in two independent datasets, and this replication suggests that these are true findings that should be explored further for their biological significance. In addition, the age at menarche association with LUMA global methylation was significant, strong, resistant to adjustment for confounders and demonstrated a dose-response effect. Despite the lack of this association in the 450K BeadChip data, the strong LUMA association calls for further investigation regarding the biology behind this finding.

In order to better understand the effect of these and other environmental exposures on DNA methylation, and to get a firmer grasp of the interplay between these exposures, DNA methylation, and disease outcomes, further investigations, with more analytic study designs, more power, and with novel epidemiological tools to strengthen causal inference are needed.

References

1 Gluckman PD, Hanson MA. Developmental Origins of Disease Paradigm: A Mechanistic and Evolutionary Perspective. *Pediatr Res* 2004; **56**: 311–7.

2 Gluckman PD, Hanson MA. Mismatch : the lifestyle diseases timebomb. Oxford, Oxford University Press, 2008.

3 Rutter M. Achievements and challenges in the biology of environmental effects. *Proc Natl Acad Sci U S A* 2012; **109 Suppl 2**: 17149–53.

4 Barouki R, Gluckman PD, Grandjean P, Hanson M, Heindel JJ. Developmental origins of non-communicable disease: implications for research and public health. *Environ Health Glob Access Sci Source* 2012; **11**: 42.

5 Gluckman PD, Hanson MA, Spencer HG, Bateson P. Environmental influences during development and their later consequences for health and disease: implications for the interpretation of empirical studies. *Proc R Soc B Biol Sci* 2005; **272**: 671–7.

6 WordNet Search - 3.1. http://wordnetweb.princeton.edu/perl/webwn?s=westernization (accessed 25 Apr2013).

7 Vineis P, Stringhini S, Porta M. Embedding research on non-communicable diseases into a life-course and trans-generational perspective. *Pers Commun*.

8 Cole TJ. The secular trend in human physical growth: a biological view. *Econ Hum Biol* 2003; **1**: 161–8.

9 Danubio ME, Sanna E. Secular changes in human biological variables in Western countries: an updated review and synthesis. *J Anthropol Sci Riv Antropol JASS Ist Ital Antropol* 2008; **86**: 91–112.

10 Health, United States, 2011: With Special Feature on Socioeconomic Status and Health. Hyattsville, MD, National Center for Health Statistics, 2012.

11 Marmot M, Allen J, Bell R, Bloomer E, Goldblatt P. WHO European review of social determinants of health and the health divide. *Lancet* 2012; **380**: 1011–29.

12 European Commission, Eurostat. Unemployment statistics - Statistics Explained. http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Unemployment_statistics (accessed 25 Apr2013).

13 Drewnowski A, Popkin BM. The Nutrition Transition: New Trends in the Global Diet. *Nutr Rev* 1997; **55**: 31–43.

14 Kakwani N, Makonnen E, van der Gaa J. Structural adjustment and living conditions in developing countries. , The World Bank, 1990.

15 Moodie R, Stuckler D, Monteiro C, *et al.* Profits and pandemics: prevention of harmful effects of tobacco, alcohol, and ultra-processed food and drink industries. *Lancet* 2013; **381**: 670–9.

16 Jahiel RI, Babor TF. Industrial epidemics, public health advocacy and the alcohol industry: lessons from other fields. *Addiction* 2007; **102**: 1335–9.

17 WHO | Global status report on noncommunicable diseases 2010. WHO. http://www.who.int/nmh/publications/ncd_report2010/en/ (accessed 23 Apr2013).

18 Basu S, McKee M, Galea G, Stuckler D. Relationship of Soft Drink Consumption to Global Overweight, Obesity, and Diabetes: A Cross-National Analysis of 75 Countries. *Am J Public Health* 2013; : e1–e7.

19 Popkin BM, Bisgrove EZ. Urbanization and nutrition in low income countries. *Food Nutr Bull* 1988; **10**: 3–23.

20 WHO | Obesity and overweight. WHO. http://www.who.int/mediacentre/factsheets/fs311/en/ (accessed 26 Apr2013).

21 Cole TJ. Establishing a standard definition for child overweight and obesity worldwide: international survey. *BMJ* 2000; **320**: 1240–1240.

22 Kuczmarski RJ FK. Increasing prevalence of overweight among us adults: The national health and nutrition examination surveys, 1960 to 1991. *JAMA* 1994; **272**: 205–11.

23 Flegal KM CM. PRevalence and trends in obesity among us adults, 1999-2008. *JAMA* 2010; **303**: 235–41.

24 Seidell J. Time Trends in Obesity: An Epidemiological Perspective. *Horm Metab Res* 2007; **29**: 155–8.

25 Stevens GA, Singh GM, Lu Y, *et al.* National, regional, and global trends in adult overweight and obesity prevalences. *Popul Health Metr* 2012; **10**: 22.

26 Von Ruesten A, Steffen A, Floegel A, *et al.* Trend in Obesity Prevalence in European Adult Cohort Populations during Follow-up since 1996 and Their Predictions to 2015. *PLoS ONE* 2011; **6**: e27455.

27 WHO | Obesity. http://www.who.int/gho/ncd/risk_factors/obesity_text/en/ (accessed 22 Apr2013).

28 Lobstein T, Frelut ML. Prevalence of overweight among children in Europe. *Obes Rev Off J Int Assoc Study Obes* 2003; **4**: 195–200.

29 Lissau I. Overweight and obesity epidemic among children. Answer from European countries. *Int J Obes Relat Metab Disord J Int Assoc Study Obes* 2004; **28 Suppl 3**: S10–15.

30 Wang Y, Lobstein T. Worldwide trends in childhood overweight and obesity. *Int J Pediatr Obes* 2006; **1**: 11–25.

31 Pan L BH. TRends in the prevalence of extreme obesity among us preschool-aged children living in low-income families, 1998-2010. *JAMA* 2012; **308**: 2563–5.

32 International Obesity Task Force (IOTF). EU Platform on Diet, Physical Activity and Health: International Obesity Task Force EU platform Briefing Paper. Brussels, International Obesity Task Force (IOTF), 2005http://www.iotf.org/media/euobesity3.pdf. (accessed 27 May2013).

33 Ahmed ML, Ong KK, Dunger DB. Childhood obesity and the timing of puberty. *Trends Endocrinol Metab TEM* 2009; **20**: 237–42.

34 Marshall WA, Tanner JM. Variations in pattern of pubertal changes in girls. *Arch Dis Child* 1969; **44**: 291–303.

35 Marshall WA, Tanner JM. Variations in the Pattern of Pubertal Changes in Boys. *Arch Dis Child* 1970; **45**: 13–23.

36 Ong KK, Ahmed ML, Dunger DB. Lessons from large population studies on timing and tempo of puberty (secular trends and relation to body size): The European trend. *Mol Cell Endocrinol* 2006; **254–255**: 8–12.

37 Liestøl K, Rosenberg M. Height, weight and menarcheal age of schoolgirls in Oslo--an update. *Ann Hum Biol* 1995; **22**: 199–205.

38 Lindgren GW, Degerfors IL, Fredriksson A, *et al.* Menarche 1990 in Stockholm schoolgirls. *Acta Paediatr Scand* 1991; **80**: 953–5.

39 Onland-Moret NC, Peeters PHM, van Gils CH, *et al.* Age at menarche in relation to adult height: the EPIC study. *Am J Epidemiol* 2005; **162**: 623–32.

40 Wyshak G, Frisch RE. Evidence for a secular trend in age of menarche. *N Engl J Med* 1982; **306**: 1033–5.

41 Himes JH. Examining the evidence for recent secular changes in the timing of puberty in US children in light of increases in the prevalence of obesity. *Mol Cell Endocrinol* 2006; **254–255**: 13–21.

42 Frisch RE, Revelle R. Height and weight at menarche and a hypothesis of critical body weights and adolescent events. *Science* 1970; **169**: 397–9.

43 Biro FM, Greenspan LC, Galvez MP. Puberty in Girls of the 21st Century. *J Pediatr Adolesc Gynecol* 2012; **25**: 289–94.

44 Matkovic V, Ilich JZ, Skugor M, *et al.* Leptin is inversely related to age at menarche in human females. *J Clin Endocrinol Metab* 1997; **82**: 3239–45.

45 Anderson GL, Limacher M, Assaf AR, *et al.* Effects of conjugated equine estrogen in postmenopausal women with hysterectomy: the Women's Health Initiative randomized controlled trial. *JAMA J Am Med Assoc* 2004; **291**: 1701–12.

46 Kaplowitz PB. Link Between Body Fat and the Timing of Puberty. *PEDIATRICS* 2008; **121**: S208–S217.

47 O'Dea J, Abraham S. Should body-mass index be used in young adolescents? *The Lancet* 1995; **345**: 657.

48 Lee JM, Appugliese D, Kaciroti N, Corwyn RF, Bradley RH, Lumeng JC. Weight Status in Young Girls and the Onset of Puberty. *Pediatrics* 2007; **119**: e624–e630.

49 Davison KK, Susman EJ, Birch LL. Percent body fat at age 5 predicts earlier pubertal development among girls at age 9. *Pediatrics* 2003; **111**: 815–21.

50 Rubin C, Maisonet M, Kieszak S, *et al.* Timing of maturation and predictors of menarche in girls enrolled in a contemporary British cohort. *Paediatr Perinat Epidemiol* 2009; **23**: 492–504.

51 Buyken AE, Karaolis-Danckert N, Remer T. Association of prepubertal body composition in healthy girls and boys with the timing of early and late pubertal markers. *Am J Clin Nutr* 2009; **89**: 221–30.

52 Parent A-S, Teilmann G, Juul A, Skakkebaek NE, Toppari J, Bourguignon J-P. The timing of normal puberty and the age limits of sexual precocity: variations around the world, secular trends, and changes after migration. *Endocr Rev* 2003; **24**: 668–93.

53 Pavelka M, Fedigan L. Menopause: a comparative life history perspective. *Yearb Phys Anthropol* 1991; : 13–38.

54 Rödström K, Bengtsson C, Milsom I, Lissner L, Sundh V, Bjoürkelund C. Evidence for a secular trend in menopausal age: a population study of women in Gothenburg. *Menopause N Y N* 2003; **10**: 538–43.

55 Nichols HB, Trentham-Dietz A, Hampton JM, *et al.* From menarche to menopause: trends among US Women born from 1912 to 1969. *Am J Epidemiol* 2006; **164**: 1003–11.

56 Takaishi M. Secular Changes in Growth of Japanese Children. *J Pediatr Endocrinol Metab* 1994; **7**. doi:10.1515/JPEM.1994.7.2.163.

57 Schmidt IM, Jørgensen MH, Michaelsen KF. Height of conscripts in Europe: is postneonatal mortality a predictor? *Ann Hum Biol* 1995; **22**: 57–67.

58 Adler NE, Ostrove JM. Socioeconomic status and health: what we know and what we don't. *Ann N Y Acad Sci* 1999; **896**: 3–15.

59 Adler NE, Boyce T, Chesney MA, *et al.* Socioeconomic status and health. The challenge of the gradient. *Am Psychol* 1994; **49**: 15–24.

60 Stringhini S, Sabia S, Shipley M, *et al.* Association of socioeconomic position with health behaviors and mortality. The Whitehall II study. *JAMA J Am Med Assoc* 2010; **303**: 1159–66.

61 Stringhini S, Dugravot A, Shipley M, *et al.* Health Behaviours, Socioeconomic Status, and Mortality: Further Analyses of the British Whitehall II and the French GAZEL Prospective Cohorts. *PLoS Med* 2011; **8**. doi:10.1371/journal.pmed.1000419.

62 Tehranifar P, Wu H-C, Fan X, *et al.* Early life socioeconomic factors and genomic DNA methylation in mid-life. *Epigenetics Off J DNA Methylation Soc* 2012; **8**.

63 Glymour MM, Avendaño M, Haas S, Berkman LF. Lifecourse social conditions and racial disparities in incidence of first stroke. *Ann Epidemiol* 2008; **18**: 904–12.

64 Galobardes B, Lynch JW, Davey Smith G. Childhood socioeconomic circumstances and cause-specific mortality in adulthood: systematic review and interpretation. *Epidemiol Rev* 2004; **26**: 7–21.

65 Lopez AD, Mathers CD, Ezzati M, Jamison DT, Murray CJ, editors. Global Burden of Disease and Risk Factors. Washington (DC), World Bank, 2006http://www.ncbi.nlm.nih.gov/books/NBK11812/ (accessed 25 Apr2013).

66 WHO | International Classification of Diseases (ICD). WHO. http://www.who.int/classifications/icd/en/ (accessed 26 Apr2013).

67 Raman SV. The Hypertensive Heart: An Integrated Understanding Informed by Imaging. *J Am Coll Cardiol* 2010; **55**: 91–6.

68 Escobar E. Hypertension and coronary heart disease. *J Hum Hypertens* 2002; **16**: S21–S26.

69 T.kuo P. Dyslipidemia and coronary artery disease. *Clin Cardiol* 1994; **17**: 519–27.

70 Lawes CMM, Vander Hoorn S, Law MR, Rodgers A. High Cholesterol. In: Comparative Quantification of Health Risks: Global and Regional Burden of Disease Attributable to Selected Major Risk Factors. Geneva, World Health Organization, 2004: 391–496.

71 Lock K, Pomerleau J, Causer L, McKee M. Low Fruit and Vegetable Consumption. In: Comparative Quantification of Health Risks: Global and Regional Burden of Disease Attributable to Selected Major Risk Factors. Geneva, World Health Organization, 2004: 597–728.

72 Zhang S, Hunter DJ, Hankinson SE, *et al.* A prospective study of folate intake and the risk of breast cancer. *JAMA J Am Med Assoc* 1999; **281**: 1632–7.

73 Olsen A, Tjønneland A, Thomsen BL, *et al.* Fruits and vegetables intake differentially affects estrogen receptor negative and positive breast cancer incidence rates. *J Nutr* 2003; **133**: 2342–7.

74 Shannon J, Ray R, Wu C, *et al.* Food and Botanical Groupings and Risk of Breast Cancer: A Case-Control Study in Shanghai, China. *Cancer Epidemiol Biomarkers Prev* 2005; **14**: 81 –90.

75 Sonestedt E, Gullberg B, Wirfält E. Both food habit change in the past and obesity status may influence the association between dietary factors and postmenopausal breast cancer. *Public Health Nutr* 2007; **10**: 769–79.

76 Trichopoulou A, Bamia C, Trichopoulos D. Anatomy of health effects of Mediterranean diet: Greek EPIC prospective cohort study. *BMJ* 2009; **338**: b2337–b2337.

77 Alshatwi AA. Breast cancer risk, dietary intake, and methylenetetrahydrofolate reductase (MTHFR)single nucleotide polymorphisms. *Food Chem Toxicol Int J Publ Br Ind Biol Res Assoc* 2010; **48**: 1881–5.

78 Maruti SS, Ulrich CM, White E. Folate and one-carbon metabolism nutrients from supplements and diet in relation to breast cancer risk. *Am J Clin Nutr* 2009; **89**: 624–33.

79 Ericson U, Sonestedt E, Gullberg B, Olsson H, Wirfält E. High folate intake is associated with lower breast cancer incidence in postmenopausal women in the Malmö Diet and Cancer cohort. *Am J Clin Nutr* 2007; **86**: 434–43.

80 Lajous M, Lazcano-Ponce E, Hernandez-Avila M, Willett W, Romieu I. Folate, vitamin B(6), and vitamin B(12) intake and the risk of breast cancer among Mexican women. *Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol* 2006; **15**: 443–8.

81 Beilby J, Ingram D, Hähnel R, Rossi E. Reduced breast cancer risk with increasing serum folate in a case-control study of the C677T genotype of the methylenetetrahydrofolate reductase gene. *Eur J Cancer Oxf Engl 1990* 2004; **40**: 1250–4.

82 Shrubsole MJ, Jin F, Dai Q, *et al.* Dietary folate intake and breast cancer risk: results from the Shanghai Breast Cancer Study. *Cancer Res* 2001; **61**: 7136–41.

83 Rohan TE, Jain MG, Howe GR, Miller AB. Dietary folate consumption and breast cancer risk. *J Natl Cancer Inst* 2000; **92**: 266–9.

84 Larsson SC, Giovannucci E, Wolk A. Folate and risk of breast cancer: a meta-analysis. *J Natl Cancer Inst* 2007; **99**: 64–76.

85 Baglietto L, Krishnan K, Severi G, *et al.* Dietary patterns and risk of breast cancer. *Br J Cancer* 2011; **104**: 524–31.

86 Sharp L, Little J, Schofield AC, *et al.* Folate and breast cancer: the role of polymorphisms in methylenetetrahydrofolate reductase (MTHFR). *Cancer Lett* 2002; **181**: 65–71.

87 Zhang FF, John EM, Knight JA, *et al.* Total energy intake and breast cancer risk in sisters: the Breast Cancer Family Registry. *Breast Cancer Res Treat* 2013; **137**: 541–51.

88 Huxley R. Excess risk of fatal coronary heart disease associated with diabetes in men and women: meta-analysis of 37 prospective cohort studies. *BMJ* 2006; **332**: 73–8.

89 Hjartåker A, Langseth H, Weiderpass E. Obesity and diabetes epidemics: cancer repercussions. *Adv Exp Med Biol* 2008; **630**: 72–93.

90 Global Status Report on Alcohol and Health 2011. Geneva, World Health Organisation, 2011.

91 Trichopoulou A, Bamia C, Lagiou P, Trichopoulos D. Conformity to traditional Mediterranean diet and breast cancer risk in the Greek EPIC (European Prospective Investigation into Cancer and Nutrition) cohort. *Am J Clin Nutr* 2010; **92**: 620–5.

92 Buckland G, Travier N, Cottet V, *et al.* Adherence to the mediterranean diet and risk of breast cancer in the EPIC cohort study. *Int J Cancer* 2012. doi:10.1002/ijc.27958.

93 Demetriou CA, Hadjisavvas A, Loizidou MA, *et al.* The mediterranean dietary pattern and breast cancer risk in Greek-Cypriot women: a case-control study. *BMC Cancer* 2012; **12**: 113.

94 Alexander JK. Obesity and Coronary Heart Disease. *Am J Med Sci* 2001; **321**: 215–24.

95 Anandacoomarasamy A, Fransen M, March L. Obesity and the musculoskeletal system. *Curr Opin Rheumatol* 2009; **21**: 71–7.

96 Chan G, Chen CT. Musculoskeletal effects of obesity. *Curr Opin Pediatr* 2009; **21**: 65–70.

97 Garcia-Closas M, Brinton LA, Lissowska J, *et al.* Established breast cancer risk factors by clinically important tumour characteristics. *Br J Cancer* 2006; **95**: 123–9.

98 Ma H, Bernstein L, Pike MC, Ursin G. Reproductive factors and breast cancer risk according to joint estrogen and progesterone receptor status: a meta-analysis of epidemiological studies. *Breast Cancer Res BCR* 2006; **8**: R43.

99 Trichopoulos D, MacMahon B, Cole P. Menopause and Breast Cancer Risk. *J Natl Cancer Inst* 1972; **48**: 605–13.

100 WHO-IARC. World Cancer Report 2008 - Breast Cancer. Lyon, France, IARC.

101 Ong KK, Northstone K, Wells JC, *et al.* Earlier Mother's Age at Menarche Predicts Rapid Infancy Growth and Childhood Obesity. *PLoS Med* 2007; **4**: e132.

102 Lakshman R, Forouhi N, Luben R, *et al.* Association between age at menarche and risk of diabetes in adults: results from the EPIC-Norfolk cohort study. *Diabetologia* 2008; **51**: 781–6.

103 Remsberg KE, Demerath EW, Schubert CM, Chumlea WC, Sun SS, Siervogel RM. Early menarche and the development of cardiovascular disease risk factors in adolescent girls: the Fels Longitudinal Study. *J Clin Endocrinol Metab* 2005; **90**: 2718–24.

104 Michaud P-A, Suris J-C, Deppen A. Gender-related psychological and behavioural correlates of pubertal timing in a national sample of Swiss adolescents. *Mol Cell Endocrinol* 2006; **254–255**: 172–8.

105 Green J, Cairns BJ, Casabonne D, Wright FL, Reeves G, Beral V. Height and cancer incidence in the Million Women Study: prospective cohort, and meta-analysis of prospective studies of height and total cancer risk. *Lancet Oncol* 2011; **12**: 785–94.

106 Case A, Paxson C. Height, Health, and Cognitive Function at Older Ages. *Am Econ Rev* 2008; **98**: 463–7.

107 Melnychuk M. Socioeconomic Status and Breast Cancer Risk: A New Methodology to Estimate the Association. Rochester, NY, Social Science Research Network, 2010http://papers.ssrn.com/abstract=1493755 (accessed 26 Apr2013).

108 Castro Martín T. Women's education and fertility: results from 26 Demographic and Health Surveys. *Stud Fam Plann* 1995; **26**: 187–202.

109 Duthie SJ. Folate and cancer: how DNA damage, repair and methylation impact on colon carcinogenesis. *J Inherit Metab Dis* 2010. doi:10.1007/s10545-010-9128-0.

110 Chen J, Gammon MD, Chan W, *et al.* One-carbon metabolism, MTHFR polymorphisms, and risk of breast cancer. *Cancer Res* 2005; **65**: 1606–14.

111 Jovanovic J, Rønneberg JA, Tost J, Kristensen V. The epigenetics of breast cancer. *Mol Oncol* 2010; **4**: 242–54.

112 Teegarden D, Romieu I, Lelièvre SA. Redefining the impact of nutrition on breast cancer incidence: is epigenetics involved? *Nutr Res Rev* 2012; **25**: 68–95.

113 Rabi T, Bishayee A. Terpenoids and breast cancer chemoprevention. *Breast Cancer Res Treat* 2009; **115**: 223–39.

114 Okuno M, Kojima S, Matsushima-Nishiwaki R, *et al.* Retinoids in cancer chemoprevention. *Curr Cancer Drug Targets* 2004; **4**: 285–98.

115 Hui C, Qi X, Qianyong Z, Xiaoli P, Jundong Z, Mantian M. Flavonoids, flavonoid subclasses and breast cancer risk: a meta-analysis of epidemiologic studies. *PloS One* 2013; **8**: e54318.

116 Humfrey CD. Phytoestrogens and human health effects: weighing up the current evidence. *Nat Toxins* 1998; **6**: 51–9.

117 Pino AM, Valladares LE, Palma MA, Mancilla AM, Yáñez M, Albala C. Dietary isoflavones affect sex hormone-binding globulin levels in postmenopausal women. *J Clin Endocrinol Metab* 2000; **85**: 2797–800.

118 Miller PE, Snyder DC. Phytochemicals and cancer risk: a review of the epidemiological evidence. *Nutr Clin Pract Off Publ Am Soc Parenter Enter Nutr* 2012; **27**: 599–612.

119 Hunter DJ, Spiegelman D, Adami H-O, *et al.* Cohort Studies of Fat Intake and the Risk of Breast Cancer — A Pooled Analysis. *N Engl J Med* 1996; **334**: 356–61.

120 Prentice RL, Caan B, Chlebowski RT, *et al.* Low-fat dietary pattern and risk of invasive breast cancer: the Women's Health Initiative Randomized Controlled Dietary Modification Trial. *JAMA J Am Med Assoc* 2006; **295**: 629–42.

121 Vucenik I, Stains JP. Obesity and cancer risk: evidence, mechanisms, and recommendations. *Ann N Y Acad Sci* 2012; **1271**: 37–43.

122 Percik R, Stumvoll M. Obesity and cancer. *Exp Clin Endocrinol Diabetes Off J Ger Soc Endocrinol Ger Diabetes Assoc* 2009; **117**: 563–6.

123 Xue F, Michels KB. Diabetes, metabolic syndrome, and breast cancer: a review of the current evidence. *Am J Clin Nutr* 2007; **86**: s823–835.

124 Cowey S, Hardy RW. The metabolic syndrome: A high-risk state for cancer? *Am J Pathol* 2006; **169**: 1505–22.

125 Kaaks R, Lukanova A, Kurzer MS. Obesity, endogenous hormones, and endometrial cancer risk: a synthetic review. *Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol* 2002; **11**: 1531–43.

126 Fenton JI, Hord NG, Lavigne JA, Perkins SN, Hursting SD. Leptin, insulin-like growth factor-1, and insulin-like growth factor-2 are mitogens in ApcMin/+ but not Apc+/+ colonic epithelial cell lines. *Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol* 2005; **14**: 1646–52.

127 Fenton SE. Endocrine-Disrupting Compounds and Mammary Gland Development: Early Exposure and Later Life Consequences. *Endocrinology* 2006; **147**: s18–24.

128 Macon MB, Fenton SE. Endocrine disruptors and the breast: early life effects and later life disease. *J Mammary Gland Biol Neoplasia* 2013; **18**: 43–61.

129 Rudel RA, Fenton SE, Ackerman JM, Euling SY, Makris SL. Environmental exposures and mammary gland development: state of the science, public health implications, and research recommendations. *Environ Health Perspect* 2011; **119**: 1053–61.

130 Renehan AG, Soerjomataram I, Leitzmann MF. Interpreting the epidemiological evidence linking obesity and cancer: A framework for population-attributable risk estimations in Europe. *Eur J Cancer Oxf Engl 1990* 2010; **46**: 2581–92.

131 Stringhini S, Polidoro S, Sacerdote C, *et al.* Epigenetic Signatures of low socio-economic status: a prospective study. *Pers Commun*.

132 Ben-Shlomo Y, Kuh D. A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives. *Int J Epidemiol* 2002; **31**: 285–93.

133 Hales CN, Barker DJ. The thrifty phenotype hypothesis. *Br Med Bull* 2001; **60**: 5–20.

134 Barker DJ, Osmond C. Low birth weight and hypertension. *BMJ* 1988; **297**: 134–5.

135 Barker DJ, Winter PD, Osmond C, Margetts B, Simmonds SJ. Weight in infancy and death from ischaemic heart disease. *Lancet* 1989; **2**: 577–80.

136 Barker DJ. The fetal and infant origins of adult disease. *BMJ* 1990; **301**: 1111.

137 Barker DJ. Fetal nutrition and cardiovascular disease in later life. *Br Med Bull* 1997; **53**: 96–108.

138 Barker DJP. Mothers, babies, and health in later life, 2nd ed. Edinburgh ; New York, Churchill Livingstone, 1998.

139 Barker DJ. The fetal origins of type 2 diabetes mellitus. *Ann Intern Med* 1999; **130**: 322–4.

140 Osmond C, Barker DJ. Fetal, infant, and childhood growth are predictors of coronary heart disease, diabetes, and hypertension in adult men and women. *Environ Health Perspect* 2000; **108**: 545–53.

141 Ravelli AC, van Der Meulen JH, Osmond C, Barker DJ, Bleker OP. Obesity at the age of 50 y in men and women exposed to famine prenatally. *Am J Clin Nutr* 1999; **70**: 811–6.

142 Heijmans BT, Tobi EW, Stein AD, *et al.* Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc Natl Acad Sci* 2008; **105**: 17046–9.

143 Zheng X, Wang Y, Ren W, *et al.* Risk of metabolic syndrome in adults exposed to the great Chinese famine during the fetal life and early childhood. *Eur J Clin Nutr* 2012; **66**: 231–6.

144 Hogg K, Price EM, Hanna CW, Robinson WP. Prenatal and perinatal environmental influences on the human fetal and placental epigenome. *Clin Pharmacol Ther* 2012; **92**: 716–26.

145 Lumey LH. Decreased birthweights in infants after maternal in utero exposure to the Dutch famine of 1944-1945. *Paediatr Perinat Epidemiol* 1992; **6**: 240–53.

146 Christensen BC, Houseman EA, Marsit CJ, *et al.* Aging and Environmental Exposures Alter Tissue-Specific DNA Methylation Dependent upon CpG Island Context. *PLoS Genet* 2009; **5**: e1000602.

147 Tobi EW, Lumey LH, Talens RP, *et al.* DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific. *Hum Mol Genet* 2009; **18**: 4046–53.

148 Esteller M. Epigenetics in cancer. *N Engl J Med* 2008; **358**: 1148–59.

149 Veeck J, Esteller M. Breast cancer epigenetics: from DNA methylation to microRNAs. *J Mammary Gland Biol Neoplasia* 2010; **15**: 5–17.

150 Vineis P, Schatzkin A, Potter JD. Models of carcinogenesis: an overview. *Carcinogenesis* 2010; **31**: 1703–9.

151 Lo P-K, Sukumar S. Epigenomics and breast cancer. *Pharmacogenomics* 2008; **9**: 1879–902.

152 Jones PA, Laird PW. Cancer epigenetics comes of age. *Nat Genet* 1999; **21**: 163–7.

153 Rodriguez-Paredes M, Esteller M. Cancer epigenetics reaches mainstream oncology. *Nat Med* 2011; : 330–9.

154 Jones PA, Baylin SB. The epigenomics of cancer. *Cell* 2007; **128**: 683–92.

155 Esteller M. Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat Rev Genet* 2007; **8**: 286–98.

156 Pfeifer GP, Besaratinia A. Mutational spectra of human cancer. *Hum Genet* 2009; **125**: 493–506.

157 De Smet C, Loriot A, Boon T. Promoter-dependent mechanism leading to selective hypomethylation within the 5’ region of gene MAGE-A1 in tumor cells. *Mol Cell Biol* 2004; **24**: 4781–90.

158 Stearns V, Zhou Q, Davidson NE. Epigenetic regulation as a new target for breast cancer therapy. *Cancer Invest* 2007; **25**: 659–65.

159 Kundu TK, Rao MR. CpG islands in chromatin organization and gene expression. *J Biochem (Tokyo)* 1999; **125**: 217–22.

160 Bird A. DNA methylation patterns and epigenetic memory. *Genes Dev* 2002; **16**: 6–21.

161 Doi A, Park I-H, Wen B, *et al.* Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat Genet* 2009; **41**: 1350–3.

162 Meissner A, Mikkelsen TS, Gu H, *et al.* Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* 2008; **454**: 766–70.

163 Hansen KD, Timp W, Bravo HC, *et al.* Increased methylation variation in epigenetic domains across cancer types. *Nat Genet* 2011; **43**: 768–75.

164 Herman JG, Baylin SB. Gene silencing in cancer in association with promoter hypermethylation. *N Engl J Med* 2003; **349**: 2042–54.

165 Costello JF, Frühwald MC, Smiraglia DJ, *et al.* Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. *Nat Genet* 2000; **24**: 132–8.

166 Ehrlich M. DNA methylation in cancer: too much, but also too little. *Oncogene* 2002; **21**: 5400–13.

167 Herceg Z, Lambert M-P, Veldhoven K van, *et al.* Towards incorporating epigenetic mechanisms into carcinogen identification and evaluation. *Carcinogenesis* 2013. doi:10.1093/carcin/bgt212.

168 Friso S, Udali S, Guarini P, *et al.* Global DNA hypomethylation in peripheral blood mononuclear cells as a biomarker of cancer risk. *Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol* 2013; **22**: 348–55.

169 Brennan K, Flanagan JM. Is there a link between genome-wide hypomethylation in blood and cancer risk? *Cancer Prev Res Phila Pa* 2012; **5**: 1345–57.

170 Cho YH, Yazici H, Wu H-C, *et al.* Aberrant promoter hypermethylation and genomic hypomethylation in tumor, adjacent normal tissues and blood from breast cancer patients. *Anticancer Res* 2010; **30**: 2489–96.

171 Choi J-Y, James SR, Link PA, *et al.* Association between global DNA hypomethylation in leukocytes and risk of breast cancer. *Carcinogenesis* 2009; **30**: 1889–97.

172 Kim YI, Pogribny IP, Salomon RN, *et al.* Exon-specific DNA hypomethylation of the p53 gene of rat colon induced by dimethylhydrazine. Modulation by dietary folate. *Am J Pathol* 1996; **149**: 1129–37.

173 Narayan A, Ji W, Zhang XY, *et al.* Hypomethylation of pericentromeric DNA in breast adenocarcinomas. *Int J Cancer J Int Cancer* 1998; **77**: 833–8.

174 Qu GZ, Grundy PE, Narayan A, Ehrlich M. Frequent hypomethylation in Wilms tumors of pericentromeric DNA in chromosomes 1 and 16. *Cancer Genet Cytogenet* 1999; **109**: 34–9.

175 Garrick D, Fiering S, Martin DI, Whitelaw E. Repeat-induced gene silencing in mammals. *Nat Genet* 1998; **18**: 56–9.

176 Nguyen C, Liang G, Nguyen TT, *et al.* Susceptibility of nonpromoter CpG islands to de novo methylation in normal and neoplastic cells. *J Natl Cancer Inst* 2001; **93**: 1465–72.

177 Vilain A, Vogt N, Dutrillaux B, Malfoy B. DNA methylation and chromosome instability in breast cancer cell lines. *FEBS Lett* 1999; **460**: 231–4.

178 Szyf M, Pakneshan P, Rabbani SA. DNA methylation and breast cancer. *Biochem Pharmacol* 2004; **68**: 1187–97.

179 Jones PA, Baylin SB. The fundamental role of epigenetic events in cancer. *Nat Rev Genet* 2002; **3**: 415–28.

180 Cui H, Cruz-Correa M, Giardiello FM, *et al.* Loss of IGF2 imprinting: a potential marker of colorectal cancer risk. *Science* 2003; **299**: 1753–5.

181 Kaneda A, Feinberg AP. Loss of imprinting of IGF2: a common epigenetic modifier of intestinal tumor risk. *Cancer Res* 2005; **65**: 11236–40.

182 Feinberg AP. Cancer epigenetics is no Mickey Mouse. *Cancer Cell* 2005; **8**: 267–8.

183 Herceg Z, Vaissière T. Epigenetic mechanisms and cancer: an interface between the environment and the genome. *Epigenetics Off J DNA Methylation Soc* 2011; **6**: 804–19.

184 Shukla V, Coumoul X, Lahusen T, *et al.* BRCA1 affects global DNA methylation through regulation of DNMT1. *Cell Res* 2010; **20**: 1201–15.

185 Girault I, Tozlu S, Lidereau R, Bièche I. Expression analysis of DNA methyltransferases 1, 3A, and 3B in sporadic breast carcinomas. *Clin Cancer Res Off J Am Assoc Cancer Res* 2003; **9**: 4415–22.

186 Szyf M. Towards a pharmacology of DNA methylation. *Trends Pharmacol Sci* 2001; **22**: 350–4.

187 Reik W. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* 2007; **447**: 425–32.

188 Weisenberger DJ, Velicescu M, Cheng JC, Gonzales FA, Liang G, Jones PA. Role of the DNA methyltransferase variant DNMT3b3 in DNA methylation. *Mol Cancer Res MCR* 2004; **2**: 62–72.

189 Roman-Gomez J, Jimenez-Velasco A, Agirre X, *et al.* Promoter hypomethylation of the LINE-1 retrotransposable elements activates sense/antisense transcription and marks the progression of chronic myeloid leukemia. *Oncogene* 2005; **24**: 7213–23.

190 Tahiliani M, Koh KP, Shen Y, *et al.* Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science* 2009; **324**: 930–5.

191 Williams K, Christensen J, Pedersen MT, *et al.* TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* 2011; **473**: 343–8.

192 Myöhänen SK, Baylin SB, Herman JG. Hypermethylation can selectively silence individual p16ink4A alleles in neoplasia. *Cancer Res* 1998; **58**: 591–3.

193 Grady WM, Willis J, Guilford PJ, *et al.* Methylation of the CDH1 promoter as the second genetic hit in hereditary diffuse gastric cancer. *Nat Genet* 2000; **26**: 16–7.

194 Wong EM, Southey MC, Fox SB, *et al.* Constitutional Methylation of the BRCA1 Promoter Is Specifically Associated with BRCA1 Mutation-Associated Pathology in Early-Onset Breast Cancer. *Cancer Prev Res (Phila Pa)* 2011; **4**: 23 –33.

195 Snell C, Krypuy M, Wong EM, Loughrey MB, Dobrovic A. BRCA1 promoter methylation in peripheral blood DNA of mutation negative familial breast cancer patients with a BRCA1 tumour phenotype. *Breast Cancer Res BCR* 2008; **10**: R12–R12.

196 Hedenfalk I, Duggan D, Chen Y, *et al.* Gene-expression profiles in hereditary breast cancer. *N Engl J Med* 2001; **344**: 539–48.

197 Van 't Veer LJ, Dai H, van de Vijver MJ, *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002; **415**: 530–6.

198 Herman JG, Merlo A, Mao L, *et al.* Inactivation of the CDKN2/p16/MTS1 gene is frequently associated with aberrant DNA methylation in all common human cancers. *Cancer Res* 1995; **55**: 4525–30.

199 Silva JM, Silva J, Sanchez A, *et al.* Tumor DNA in plasma at diagnosis of breast cancer patients is a valuable predictor of disease-free survival. *Clin Cancer Res Off J Am Assoc Cancer Res* 2002; **8**: 3761–6.

200 Esteller M, Toyota M, Sanchez-Cespedes M, *et al.* Inactivation of the DNA repair gene O6-methylguanine-DNA methyltransferase by promoter hypermethylation is associated with G to A mutations in K-ras in colorectal tumorigenesis. *Cancer Res* 2000; **60**: 2368–71.

201 Herman JG, Civin CI, Issa JP, Collector MI, Sharkis SJ, Baylin SB. Distinct patterns of inactivation of p15INK4B and p16INK4A characterize the major types of hematological malignancies. *Cancer Res* 1997; **57**: 837–41.

202 Niwa Y, Oyama T, Nakajima T. BRCA1 expression status in relation to DNA methylation of the BRCA1 promoter region in sporadic breast cancers. *Jpn J Cancer Res Gann* 2000; **91**: 519–26.

203 Rice JC, Ozcelik H, Maxeiner P, Andrulis I, Futscher BW. Methylation of the BRCA1 promoter is associated with decreased BRCA1 mRNA levels in clinical breast cancer specimens. *Carcinogenesis* 2000; **21**: 1761–5.

204 Murata H, Khattar NH, Kang Y, Gu L, Li G-M. Genetic and epigenetic modification of mismatch repair genes hMSH2 and hMLH1 in sporadic breast cancer with microsatellite instability. *Oncogene* 2002; **21**: 5696–703.

205 Klose RJ, Bird AP. Genomic DNA methylation: the mark and its mediators. *Trends Biochem Sci* 2006; **31**: 89–97.

206 Clouaire T, Stancheva I. Methyl-CpG binding proteins: specialized transcriptional repressors or structural components of chromatin? *Cell Mol Life Sci CMLS* 2008; **65**: 1509–22.

207 Portela A, Esteller M. Epigenetic modifications and human disease. *Nat Biotech* 2010; **28**: 1057–68.

208 Nan X, Ng HH, Johnson CA, *et al.* Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* 1998; **393**: 386–9.

209 Jones PL, Veenstra GJ, Wade PA, *et al.* Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet* 1998; **19**: 187–91.

210 Fuks F, Hurd PJ, Wolf D, Nan X, Bird AP, Kouzarides T. The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. *J Biol Chem* 2003; **278**: 4035–40.

211 Kimura H, Shiota K. Methyl-CpG-binding protein, MeCP2, is a target molecule for maintenance DNA methyltransferase, Dnmt1. *J Biol Chem* 2003; **278**: 4806–12.

212 Sarraf SA, Stancheva I. Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. *Mol Cell* 2004; **15**: 595–605.

213 Robertson KD, Ait-Si-Ali S, Yokochi T, Wade PA, Jones PL, Wolffe AP. DNMT1 forms a complex with Rb, E2F1 and HDAC1 and represses transcription from E2F-responsive promoters. *Nat Genet* 2000; **25**: 338–42.

214 Rountree MR, Bachman KE, Baylin SB. DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci. *Nat Genet* 2000; **25**: 269–77.

215 Fuks F, Hurd PJ, Deplus R, Kouzarides T. The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. *Nucleic Acids Res* 2003; **31**: 2305–12.

216 Bachman KE, Park BH, Rhee I, *et al.* Histone modifications and silencing prior to DNA methylation of a tumor suppressor gene. *Cancer Cell* 2003; **3**: 89–95.

217 Tamaru H, Selker EU. A histone H3 methyltransferase controls DNA methylation in Neurospora crassa. *Nature* 2001; **414**: 277–83.

218 Martin C, Zhang Y. The diverse functions of histone lysine methylation. *Nat Rev Mol Cell Biol* 2005; **6**: 838–49.

219 Keshet I, Schlesinger Y, Farkash S, *et al.* Evidence for an instructive mechanism of de novo methylation in cancer cells. *Nat Genet* 2006; **38**: 149–53.

220 Schlesinger Y, Straussman R, Keshet I, *et al.* Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. *Nat Genet* 2007; **39**: 232–6.

221 Lachner M, O'Carroll D, Rea S, Mechtler K, Jenuwein T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* 2001; **410**: 116–20.

222 Viré E, Brenner C, Deplus R, *et al.* The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* 2006; **439**: 871–4.

223 Zhao Q, Rank G, Tan YT, *et al.* PRMT5-mediated methylation of histone H4R3 recruits DNMT3A, coupling histone and DNA methylation in gene silencing. *Nat Struct Mol Biol* 2009; **16**: 304–11.

224 Estève P-O, Chin HG, Benner J, *et al.* Regulation of DNMT1 stability through SET7-mediated lysine methylation in mammalian cells. *Proc Natl Acad Sci* 2009; **106**: 5076 –5081.

225 Ooi SKT, Qiu C, Bernstein E, *et al.* DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* 2007; **448**: 714–7.

226 Novak P, Jensen T, Oshiro MM, *et al.* Epigenetic Inactivation of the HOXA Gene Cluster in Breast Cancer. *Cancer Res* 2006; **66**: 10664 –10670.

227 Yebra MJ, Bhagwat AS. A cytosine methyltransferase converts 5-methylcytosine in DNA to thymine. *Biochemistry (Mosc)* 1995; **34**: 14752–7.

228 Choi SW, Mason JB. Folate and carcinogenesis: an integrated scheme. *J Nutr* 2000; **130**: 129–32.

229 Jones S, Zhang X, Parsons DW, *et al.* Core Signaling Pathways in Human Pancreatic Cancers Revealed by Global Genomic Analyses. *Science* 2008; **321**: 1801 –1806.

230 Burbee DG, Forgacs E, Zöchbauer-Müller S, *et al.* Epigenetic inactivation of RASSF1A in lung and breast cancers and malignant phenotype suppression. *J Natl Cancer Inst* 2001; **93**: 691–9.

231 Dammann R, Li C, Yoon JH, Chin PL, Bates S, Pfeifer GP. Epigenetic inactivation of a RAS association domain family protein from the lung tumour suppressor locus 3p21.3. *Nat Genet* 2000; **25**: 315–9.

232 Wales MM, Biel MA, el Deiry W, *et al.* p53 activates expression of HIC-1, a new candidate tumour suppressor gene on 17p13.3. *Nat Med* 1995; **1**: 570–7.

233 Esteller M, Corn PG, Baylin SB, Herman JG. A gene hypermethylation profile of human cancer. *Cancer Res* 2001; **61**: 3225–9.

234 Feltus FA, Lee EK, Costello JF, Plass C, Vertino PM. Predicting aberrant CpG island methylation. *Proc Natl Acad Sci U S A* 2003; **100**: 12253–8.

235 Esteller M. Epigenetic gene silencing in cancer: the DNA hypermethylome. *Hum Mol Genet* 2007; **16 Spec No 1**: R50–59.

236 Feltus FA, Lee EK, Costello JF, Plass C, Vertino PM. DNA motifs associated with aberrant CpG island methylation. *Genomics* 2006; **87**: 572–9.

237 Graff JR, Herman JG, Myöhänen S, Baylin SB, Vertino PM. Mapping patterns of CpG island methylation in normal and neoplastic cells implicates both upstream and downstream regions in de novo methylation. *J Biol Chem* 1997; **272**: 22322–9.

238 Wilson AS, Power BE, Molloy PL. DNA hypomethylation and human diseases. *Biochim Biophys Acta* 2007; **1775**: 138–62.

239 Weber J, Salgaller M, Samid D, *et al.* Expression of the MAGE-1 tumor antigen is up-regulated by the demethylating agent 5-aza-2'-deoxycytidine. *Cancer Res* 1994; **54**: 1766–71.

240 Fisher JL, Field CL, Zhou H, Harris TL, Henderson MA, Choong PF. Urokinase plasminogen activator system gene expression is increased in human breast carcinoma and its bone metastases--a comparison of normal breast tissue, non-invasive and invasive carcinoma and osseous metastases. *Breast Cancer Res Treat* 2000; **61**: 1–12.

241 Guo Y, Pakneshan P, Gladu J, Slack A, Szyf M, Rabbani SA. Regulation of DNA methylation in human breast cancer. Effect on the urokinase-type plasminogen activator gene production and tumor invasion. *J Biol Chem* 2002; **277**: 41571–9.

242 Lu A, Gupta A, Li C, *et al.* Molecular mechanisms for aberrant expression of the human breast cancer specific gene 1 in breast cancer cells: control of transcription by DNA methylation and intronic sequences. *Oncogene* 2001; **20**: 5173–85.

243 Gupta A, Godwin AK, Vanderveer L, Lu A, Liu J. Hypomethylation of the synuclein gamma gene CpG island promotes its aberrant expression in breast carcinoma and ovarian carcinoma. *Cancer Res* 2003; **63**: 664–73.

244 Sharma G, Mirza S, Parshad R, *et al.* CpG hypomethylation of MDR1 gene in tumor and serum of invasive ductal breast carcinoma patients. *Clin Biochem* 2010; **43**: 373–9.

245 Leick MB, Shoff CJ, Wang EC, Congress JL, Gallicano GI. Loss of imprinting of IGF2 and the epigenetic progenitor model of cancer. *Am J Stem Cells* 2011; **1**: 59–74.

246 Ito Y, Koessler T, Ibrahim AEK, *et al.* Somatically acquired hypomethylation of IGF2 in breast and colorectal cancer. *Hum Mol Genet* 2008; **17**: 2633 –2643.

247 Ogawa O, Eccles MR, Szeto J, *et al.* Relaxation of insulin-like growth factor II gene imprinting implicated in Wilms' tumour. *Nature* 1993; **362**: 749–51.

248 Watt PM, Kumar R, Kees UR. Promoter demethylation accompanies reactivation of the HOX11 proto-oncogene in leukemia. *Genes Chromosomes Cancer* 2000; **29**: 371–7.

249 Irizarry RA, Ladd-Acosta C, Wen B, *et al.* The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet* 2009; **41**: 178–86.

250 Ji H, Ehrlich LIR, Seita J, *et al.* Comprehensive methylome map of lineage commitment from haematopoietic progenitors. *Nature* 2010; **467**: 338–42.

251 Cheng ASL, Culhane AC, Chan MWY, *et al.* Epithelial progeny of estrogen-exposed breast progenitor cells display a cancer-like methylome. *Cancer Res* 2008; **68**: 1786–96.

252 Bloushtain-Qimron N, Yao J, Snyder EL, *et al.* Cell type-specific DNA methylation patterns in the human breast. *Proc Natl Acad Sci U S A* 2008; **105**: 14076–81.

253 Shipitsin M, Campbell LL, Argani P, *et al.* Molecular definition of breast tumor heterogeneity. *Cancer Cell* 2007; **11**: 259–73.

254 Leu Y-W, Yan PS, Fan M, *et al.* Loss of estrogen receptor signaling triggers epigenetic silencing of downstream targets in breast cancer. *Cancer Res* 2004; **64**: 8184–92.

255 Fan M, Yan PS, Hartman-Frey C, *et al.* Diverse gene expression and DNA methylation profiles correlate with differential adaptation of breast cancer cells to the antiestrogens tamoxifen and fulvestrant. *Cancer Res* 2006; **66**: 11954–66.

256 Iorns E, Turner NC, Elliott R, *et al.* Identification of CDK10 as an important determinant of resistance to endocrine therapy for breast cancer. *Cancer Cell* 2008; **13**: 91–104.

257 Ley TJ, Ding L, Walter MJ, *et al.* DNMT3A mutations in acute myeloid leukemia. *N Engl J Med* 2010; **363**: 2424–33.

258 Garzon R, Liu S, Fabbri M, *et al.* MicroRNA-29b induces global DNA hypomethylation and tumor suppressor gene reexpression in acute myeloid leukemia by targeting directly DNMT3A and 3B and indirectly DNMT1. *Blood* 2009; **113**: 6411–8.

259 Martens JW, Margossian AL, Schmitt M, Foekens J, Harbeck N. DNA methylation as a biomarker in breast cancer. *Future Oncol Lond Engl* 2009; **5**: 1245–56.

260 Berg WA, Gutierrez L, NessAiver MS, *et al.* Diagnostic Accuracy of Mammography, Clinical Examination, US, and MR Imaging in Preoperative Assessment of Breast Cancer1. *Radiology* 2004; **233**: 830 –849.

261 Evans A. Hormone replacement therapy and mammographic screening. *Clin Radiol* 2002; **57**: 563–4.

262 Zhi H, Ou B, Luo B-M, Feng X, Wen Y-L, Yang H-Y. Comparison of ultrasound elastography, mammography, and sonography in the diagnosis of solid breast lesions. *J Ultrasound Med Off J Am Inst Ultrasound Med* 2007; **26**: 807–15.

263 Maskarinec G, Meng L, Ursin G. Ethnic differences in mammographic densities. *Int J Epidemiol* 2001; **30**: 959–65.

264 Brooks J, Cairns P, Zeleniuch-Jacquotte A. Promoter methylation and the detection of breast cancer. *Cancer Causes Control CCC* 2009; **20**: 1539–50.

265 Elmore JG, Barton MB, Moceri VM, Polk S, Arena PJ, Fletcher SW. Ten-year risk of false positive screening mammograms and clinical breast examinations. *N Engl J Med* 1998; **338**: 1089–96.

266 Ernster VL, Barclay J. Increases in ductal carcinoma in situ (DCIS) of the breast in relation to mammography: a dilemma. *J Natl Cancer Inst Monogr* 1997; : 151–6.

267 Jeronimo C, Monteiro P, Henrique R, *et al.* Quantitative hypermethylation of a small panel of genes augments the diagnostic accuracy in fine-needle aspirate washings of breast lesions. *Breast Cancer Res Treat* 2008; **109**: 27–34.

268 Dulaimi E, Hillinck J, Ibanez de Caceres I, Al-Saleem T, Cairns P. Tumor suppressor gene promoter hypermethylation in serum of breast cancer patients. *Clin Cancer Res Off J Am Assoc Cancer Res* 2004; **10**: 6189–93.

269 Parrella P, Poeta ML, Gallo AP, *et al.* Nonrandom distribution of aberrant promoter methylation of cancer-related genes in sporadic breast tumors. *Clin Cancer Res Off J Am Assoc Cancer Res* 2004; **10**: 5349–54.

270 Lewis CM, Cler LR, Bu D-W, *et al.* Promoter hypermethylation in benign breast epithelium in relation to predicted breast cancer risk. *Clin Cancer Res Off J Am Assoc Cancer Res* 2005; **11**: 166–72.

271 Pu RT, Laitala LE, Alli PM, Fackler MJ, Sukumar S, Clark DP. Methylation profiling of benign and malignant breast lesions and its application to cytopathology. *Mod Pathol Off J U S Can Acad Pathol Inc* 2003; **16**: 1095–101.

272 Evron E, Dooley WC, Umbricht CB, *et al.* Detection of breast cancer cells in ductal lavage fluid by methylation-specific PCR. *Lancet* 2001; **357**: 1335–6.

273 Fackler MJ, McVeigh M, Evron E, *et al.* DNA methylation ofRASSF1A, HIN-1, RAR-?, Cyclin D2 andTwist inin situ and invasive lobular breast carcinoma. *Int J Cancer* 2003; **107**: 970–5.

274 Eads CA, Danenberg KD, Kawakami K, *et al.* MethyLight: a high-throughput assay to measure DNA methylation. *Nucleic Acids Res* 2000; **28**: E32.

275 Bae YK, Brown A, Garrett E, *et al.* Hypermethylation in histologically distinct classes of breast cancer. *Clin Cancer Res Off J Am Assoc Cancer Res* 2004; **10**: 5998–6005.

276 Kelly TK, De Carvalho DD, Jones PA. Epigenetic modifications as therapeutic targets. *Nat Biotech* 2010; **28**: 1069–78.

277 Wong EM, Southey MC, Fox SB, *et al.* Constitutional Methylation of the BRCA1 Promoter Is Specifically Associated with BRCA1 Mutation-Associated Pathology in Early-Onset Breast Cancer. *Cancer Prev Res (Phila Pa)* 2011; **4**: 23 –33.

278 Widschwendter M, Apostolidou S, Raum E, *et al.* Epigenotyping in Peripheral Blood Cell DNA and Breast Cancer Risk: A Proof of Principle Study. *PLoS ONE* 2008; **3**: e2656.

279 Langevin SM, Koestler DC, Christensen BC, *et al.* Peripheral blood DNA methylation profiles are indicative of head and neck squamous cell carcinoma: an epigenome-wide association study. *Epigenetics Off J DNA Methylation Soc* 2012; **7**: 291–9.

280 Teschendorff AE, Menon U, Gentry-Maharaj A, *et al.* An Epigenetic Signature in Peripheral Blood Predicts Active Ovarian Cancer. *PLoS ONE* 2009; **4**: e8274.

281 Brennan K, Garcia-Closas M, Orr N, *et al.* Intragenic ATM methylation in peripheral blood DNA as a biomarker of breast cancer risk. *Cancer Res* 2012; **72**: 2304–13.

282 Flanagan JM, Munoz-Alegre M, Henderson S, *et al.* Gene-body hypermethylation of ATM in peripheral blood DNA of bilateral breast cancer patients. *Hum Mol Genet* 2009; **18**: 1332–42.

283 Widschwendter M, Menon U. Circulating methylated DNA: a new generation of tumor markers. *Clin Cancer Res Off J Am Assoc Cancer Res* 2006; **12**: 7205–8.

284 Müller HM, Widschwendter A, Fiegl H, *et al.* DNA methylation in serum of breast cancer patients: an independent prognostic marker. *Cancer Res* 2003; **63**: 7641–5.

285 Jahr S, Hentze H, Englisch S, *et al.* DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. *Cancer Res* 2001; **61**: 1659–65.

286 Hoque MO, Feng Q, Toure P, *et al.* Detection of aberrant methylation of four genes in plasma DNA for the detection of breast cancer. *J Clin Oncol Off J Am Soc Clin Oncol* 2006; **24**: 4262–9.

287 Herceg Z. Epigenetics and cancer: towards an evaluation of the impact of environmental and dietary factors. *Mutagenesis* 2007; **22**: 91 –103.

288 Jacob RA, Gretz DM, Taylor PC, *et al.* Moderate folate depletion increases plasma homocysteine and decreases lymphocyte DNA methylation in postmenopausal women. *J Nutr* 1998; **128**: 1204–12.

289 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.

290 Choi S-W, Mason JB. Folate status: effects on pathways of colorectal carcinogenesis. *J Nutr* 2002; **132**: 2413S–2418S.

291 Shenker NS, Polidoro S, van Veldhoven K, *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* 2012. doi:10.1093/hmg/dds488.

292 Zhu Z-Z, Hou L, Bollati V, *et al.* Predictors of global methylation levels in blood DNA of healthy subjects: a combined analysis. *Int J Epidemiol* 2012; **41**: 126–39.

293 Koestler DC, Marsit CJ, Christensen BC, *et al.* Peripheral blood immune cell methylation profiles are associated with nonhematopoietic cancers. *Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol* 2012; **21**: 1293–302.

294 Ottaviano YL, Issa JP, Parl FF, Smith HS, Baylin SB, Davidson NE. Methylation of the estrogen receptor gene CpG island marks loss of estrogen receptor expression in human breast cancer cells. *Cancer Res* 1994; **54**: 2552–5.

295 Lapidus RG, Nass SJ, Butash KA, *et al.* Mapping of ER gene CpG island methylation-specific polymerase chain reaction. *Cancer Res* 1998; **58**: 2515–9.

296 Billam M, Witt AE, Davidson NE. The silent estrogen receptor: Can we make it speak? *Cancer Biol Ther* 2009; **8**.http://www.ncbi.nlm.nih.gov/pubmed/19411863 (accessed 6 Jun2011).

297 Balch C, Yan P, Craft T, *et al.* Antimitogenic and chemosensitizing effects of the methylation inhibitor zebularine in ovarian cancer. *Mol Cancer Ther* 2005; **4**: 1505 –1514.

298 Teodoridis JM, Strathdee G, Plumb JA, Brown R. CpG-island methylation and epigenetic control of resistance to chemotherapy. *Biochem Soc Trans* 2004; **32**: 916–7.

299 Teodoridis JM, Hall J, Marsh S, *et al.* CpG island methylation of DNA damage response genes in advanced ovarian cancer. *Cancer Res* 2005; **65**: 8961–7.

300 Plumb JA, Strathdee G, Sludden J, Kaye SB, Brown R. Reversal of drug resistance in human tumor xenografts by 2'-deoxy-5-azacytidine-induced demethylation of the hMLH1 gene promoter. *Cancer Res* 2000; **60**: 6039–44.

301 Veeck J, Ropero S, Setien F, *et al.* BRCA1 CpG Island Hypermethylation Predicts Sensitivity to Poly(Adenosine Diphosphate)- Ribose Polymerase Inhibitors. *J Clin Oncol* 2010; **28**: e563 –e564.

302 Dai W, Zeller C, Masrour N, Siddiqui N, Paul J, Brown R. Promoter CpG Island Methylation of Genes in Key Cancer Pathways Associates with Clinical Outcome in High Grade Serous Ovarian Cancer. *Clin Cancer Res Off J Am Assoc Cancer Res* 2013. doi:10.1158/1078-0432.CCR-13-1217.

303 Zeller C, Brown R. Therapeutic modulation of epigenetic drivers of drug resistance in ovarian cancer. *Ther Adv Med Oncol* 2010; **2**: 319–29.

304 Hellebrekers DMEI, Griffioen AW, van Engeland M. Dual targeting of epigenetic therapy in cancer. *Biochim Biophys Acta* 2007; **1775**: 76–91.

305 Jones PA, Taylor SM. Cellular differentiation, cytidine analogs and DNA methylation. *Cell* 1980; **20**: 85–93.

306 Villar-Garea A, Fraga MF, Espada J, Esteller M. Procaine is a DNA-demethylating agent with growth-inhibitory effects in human cancer cells. *Cancer Res* 2003; **63**: 4984–9.

307 Segura-Pacheco B, Trejo-Becerril C, Perez-Cardenas E, *et al.* Reactivation of tumor suppressor genes by the cardiovascular drugs hydralazine and procainamide and their potential use in cancer therapy. *Clin Cancer Res Off J Am Assoc Cancer Res* 2003; **9**: 1596–603.

308 Fang MZ, Wang Y, Ai N, *et al.* Tea polyphenol (-)-epigallocatechin-3-gallate inhibits DNA methyltransferase and reactivates methylation-silenced genes in cancer cell lines. *Cancer Res* 2003; **63**: 7563–70.

309 Brueckner B, Garcia Boy R, Siedlecki P, *et al.* Epigenetic reactivation of tumor suppressor genes by a novel small-molecule inhibitor of human DNA methyltransferases. *Cancer Res* 2005; **65**: 6305–11.

310 French HJ, Attenborough R, Hardy K, Shannon MF, Williams RBH. Interindividual variation in epigenomic phenomena in humans. *Mamm Genome* 2009; **20**: 604–11.

311 Relton CL, Davey Smith G. Epigenetic Epidemiology of Common Complex Disease: Prospects for Prediction, Prevention, and Treatment. *PLoS Med* 2010; **7**. doi:10.1371/journal.pmed.1000356.

312 Borghol N, Suderman M, McArdle W, *et al.* Associations with early-life socio-economic position in adult DNA methylation. *Int J Epidemiol* 2012; **41**: 62–74.

313 McGuinness D, McGlynn LM, Johnson PCD, *et al.* Socio-economic status is associated with epigenetic differences in the pSoBid cohort. *Int J Epidemiol* 2012; **41**: 151–60.

314 Lam LL, Emberly E, Fraser HB, *et al.* Factors underlying variable DNA methylation in a human community cohort. *Proc Natl Acad Sci U S A* 2012; **109 Suppl 2**: 17253–60.

315 Miller GE, Chen E, Fok AK, *et al.* Low early-life social class leaves a biological residue manifested by decreased glucocorticoid and increased proinflammatory signaling. *Proc Natl Acad Sci* 2009; **106**: 14716–21.

316 Lumey LH, Terry MB, Delgado-Cruzata L, *et al.* Adult global DNA methylation in relation to pre-natal nutrition. *Int J Epidemiol* 2011. doi:10.1093/ije/dyr137.

317 Terry MB, Ferris JS, Pilsner R, *et al.* Genomic DNA Methylation among Women in a Multiethnic New York City Birth Cohort. *Cancer Epidemiol Biomarkers Prev* 2008; **17**: 2306 –2310.

318 Subramanyam MA, Diez-Roux AV, Pilsner JR, *et al.* Social factors and leukocyte DNA methylation of repetitive sequences: the multi-ethnic study of atherosclerosis. *PloS One* 2013; **8**: e54018.

319 Xu X, Gammon MD, Zhang H, *et al.* Polymorphisms of one-carbon-metabolizing genes and risk of breast cancer in a population-based study. *Carcinogenesis* 2007; **28**: 1504–9.

320 Pufulete M, Al-Ghnaniem R, Leather AJM, *et al.* Folate status, genomic DNA hypomethylation, and risk of colorectal adenoma and cancer: a case control study. *Gastroenterology* 2003; **124**: 1240–8.

321 Friso S, Choi S-W. Gene-Nutrient Interactions and DNA Methylation. *J Nutr* 2002; **132**: 2382S –2387S.

322 Fenech M, Aitken C, Rinaldi J. Folate, vitamin B12, homocysteine status and DNA damage in young Australian adults. *Carcinogenesis* 1998; **19**: 1163–71.

323 Ingrosso D, Cimmino A, Perna AF, *et al.* Folate treatment and unbalanced methylation and changes of allelic expression induced by hyperhomocysteinaemia in patients with uraemia. *The Lancet* 2003; **361**: 1693–9.

324 Kim Y-I. Folate and colorectal cancer: an evidence-based critical review. *Mol Nutr Food Res* 2007; **51**: 267–92.

325 Kim YI. Folate and carcinogenesis: evidence, mechanisms, and implications. *J Nutr Biochem* 1999; **10**: 66–88.

326 Wainfan E, Dizik M, Stender M, Christman JK. Rapid appearance of hypomethylated DNA in livers of rats fed cancer-promoting, methyl-deficient diets. *Cancer Res* 1989; **49**: 4094–7.

327 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**: 1942–7.

328 Kim YI, Pogribny IP, Salomon RN, *et al.* Exon-specific DNA hypomethylation of the p53 gene of rat colon induced by dimethylhydrazine. Modulation by dietary folate. *Am J Pathol* 1996; **149**: 1129–37.

329 Kim YI, Pogribny IP, Basnakian AG, *et al.* Folate deficiency in rats induces DNA strand breaks and hypomethylation within the p53 tumor suppressor gene. *Am J Clin Nutr* 1997; **65**: 46–52.

330 Shen JC, Rideout WM, Jones PA. High frequency mutagenesis by a DNA methyltransferase. *Cell* 1992; **71**: 1073–80.

331 Pogribny IP, Basnakian AG, Miller BJ, Lopatina NG, Poirier LA, James SJ. Breaks in genomic DNA and within the p53 gene are associated with hypomethylation in livers of folate/methyl-deficient rats. *Cancer Res* 1995; **55**: 1894–901.

332 Blount BC, Ames BN. Analysis of uracil in DNA by gas chromatography-mass spectrometry. *Anal Biochem* 1994; **219**: 195–200.

333 Blount BC, Mack MM, Wehr CM, *et al.* Folate deficiency causes uracil misincorporation into human DNA and chromosome breakage: implications for cancer and neuronal damage. *Proc Natl Acad Sci U S A* 1997; **94**: 3290–5.

334 Lopatina NG, Vanyushin BF, Cronin GM, Poirier LA. Elevated expression and altered pattern of activity of DNA methyltransferase in liver tumors of rats fed methyl-deficient diets. *Carcinogenesis* 1998; **19**: 1777–81.

335 James SJ, Pogribny IP, Pogribna M, Miller BJ, Jernigan S, Melnyk S. Mechanisms of DNA damage, DNA hypomethylation, and tumor progression in the folate/methyl-deficient rat model of hepatocarcinogenesis. *J Nutr* 2003; **133**: 3740S–3747S.

336 Mason JB, Choi S-W. Effects of alcohol on folate metabolism: implications for carcinogenesis. *Alcohol Fayettev N* 2005; **35**: 235–41.

337 Manari AP, Preedy VR, Peters TJ. Nutritional intake of hazardous drinkers and dependent alcoholics in the UK. *Addict Biol* 2003; **8**: 201–10.

338 Halsted CH, Robles EA, Mezey E. Decreased jejunal uptake of labeled folic acid (3 H-PGA) in alcoholic patients: roles of alcohol and nutrition. *N Engl J Med* 1971; **285**: 701–6.

339 Halsted CH, Robles EA, Mezey E. Intestinal malabsorption in folate-deficient alcoholics. *Gastroenterology* 1973; **64**: 526–32.

340 McMartin KE, Collins TD, Eisenga BH, Fortney T, Bates WR, Bairnsfather L. Effects of chronic ethanol and diet treatment on urinary folate excretion and development of folate deficiency in the rat. *J Nutr* 1989; **119**: 1490–7.

341 Shaw S, Jayatilleke E, Herbert V, Colman N. Cleavage of folates during ethanol metabolism. Role of acetaldehyde/xanthine oxidase-generated superoxide. *Biochem J* 1989; **257**: 277–80.

342 Halsted CH, Villanueva JA, Devlin AM, *et al.* Folate deficiency disturbs hepatic methionine metabolism and promotes liver injury in the ethanol-fed micropig. *Proc Natl Acad Sci* 2002; **99**: 10072 –10077.

343 Barak AJ, Beckenhauer HC, Tuma DJ, Badakhsh S. Effects of prolonged ethanol feeding on methionine metabolism in rat liver. *Biochem Cell Biol Biochim Biol Cell* 1987; **65**: 230–3.

344 Laufer EM, Hartman TJ, Baer DJ, *et al.* Effects of moderate alcohol consumption on folate and vitamin B(12) status in postmenopausal women. *Eur J Clin Nutr* 2004; **58**: 1518–24.

345 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 –29323.

346 Tao MH, Marian C, Shields PG, *et al.* Alcohol consumption in relation to aberrant DNA methylation in breast tumors. *Alcohol Fayettev N* 2010. doi:10.1016/j.alcohol.2010.11.006.

347 Meng Q, Gao B, Goldberg ID, Rosen EM, Fan S. Stimulation of cell invasion and migration by alcohol in breast cancer cells. *Biochem Biophys Res Commun* 2000; **273**: 448–53.

348 Widschwendter M, Jones PA. DNA methylation and breast carcinogenesis. *Oncogene* 2002; **21**: 5462–82.

349 Zhu K, Davidson NE, Hunter S, *et al.* Methyl-group dietary intake and risk of breast cancer among African-American women: a case-control study by methylation status of the estrogen receptor alpha genes. *Cancer Causes Control CCC* 2003; **14**: 827–36.

350 Mills KD, Ferguson DO, Alt FW. The role of DNA breaks in genomic instability and tumorigenesis. *Immunol Rev* 2003; **194**: 77–95.

351 Villanueva JA, Halsted CH. Hepatic transmethylation reactions in micropigs with alcoholic liver disease. *Hepatol Baltim Md* 2004; **39**: 1303–10.

352 Fell D, Selhub J. Disruption of thymidylate synthesis and glycine-serine interconversion by l-methionine and l-homocystine in Raji cells. *Biochim Biophys Acta BBA - Gen Subj* 1990; **1033**: 80–4.

353 Barak AJ, Beckenhauer HC, Tuma DJ. Betaine effects on hepatic methionine metabolism elicited by short-term ethanol feeding. *Alcohol Fayettev N* 1996; **13**: 483–6.

354 Barak AJ, Beckenhauer HC, Mailliard ME, Kharbanda KK, Tuma DJ. Betaine lowers elevated s-adenosylhomocysteine levels in hepatocytes from ethanol-fed rats. *J Nutr* 2003; **133**: 2845–8.

355 Ji C, Kaplowitz N. Betaine decreases hyperhomocysteinemia, endoplasmic reticulum stress, and liver injury in alcohol-fed mice. *Gastroenterology* 2003; **124**: 1488–99.

356 Lieber CS, Leo MA, Mak KM, DeCarli LM, Sato S. Choline fails to prevent liver fibrosis in ethanol-fed baboons but causes toxicity. *Hepatol Baltim Md* 1985; **5**: 561–72.

357 Michels KB, Harris HR, Barault L. Birthweight, maternal weight trajectories and global DNA methylation of LINE-1 repetitive elements. *PloS One* 2011; **6**: e25254.

358 Kuehnen P, Mischke M, Wiegand S, *et al.* An Alu element-associated hypermethylation variant of the POMC gene is associated with childhood obesity. *PLoS Genet* 2012; **8**: e1002543.

359 Wang X, Zhu H, Snieder H, *et al.* Obesity related methylation changes in DNA of peripheral blood leukocytes. *BMC Med* 2010; **8**: 87.

360 Huang R-C, Galati JC, Burrows S, *et al.* DNA methylation of the IGF2/H19 imprinting control region and adiposity distribution in young adults. *Clin Epigenetics* 2012; **4**: 21.

361 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**: 1180–94.

362 Stepanow S, Reichwald K, Huse K, *et al.* Allele-specific, age-dependent and BMI-associated DNA methylation of human MCHR1. *PloS One* 2011; **6**: e17711.

363 Piyathilake C, Badiga S, Johanning G, Alvarez R, Partridge E. Predictors and Health Consequences of Epigenetic Changes Associated with Excess Body Weight in Women of Child-bearing Age. *Cancer Epidemiol Biomarkers Prev* 2011; **20**: 719–719.

364 Feinberg AP, Irizarry RA, Fradin D, *et al.* Personalized epigenomic signatures that are stable over time and covary with body mass index. *Sci Transl Med* 2010; **2**: 49ra67.

365 Zhang FF, Cardarelli R, Carroll J, *et al.* Significant differences in global genomic DNA methylation by gender and race/ethnicity in peripheral blood. *Epigenetics Off J DNA Methylation Soc* 2011; **6**: 623–9.

366 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.

367 Tao M-H, Mason JB, Marian C, *et al.* Promoter methylation of E-cadherin, p16, and RAR-β(2) genes in breast tumors and dietary intake of nutrients important in one-carbon metabolism. *Nutr Cancer* 2011; **63**: 1143–50.

368 Peters I, Vaske B, Albrecht K, Kuczyk MA, Jonas U, Serth J. Adiposity and age are statistically related to enhanced RASSF1A tumor suppressor gene promoter methylation in normal autopsy kidney tissue. *Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol* 2007; **16**: 2526–32.

369 Slattery ML, Curtin K, Sweeney C, *et al.* Diet and lifestyle factor associations with CpG island methylator phenotype and BRAF mutations in colon cancer. *Int J Cancer J Int Cancer* 2007; **120**: 656–63.

370 Perkins E, Murphy SK, Murtha AP, *et al.* Insulin-like growth factor 2/H19 methylation at birth and risk of overweight and obesity in children. *J Pediatr* 2012; **161**: 31–9.

371 Godfrey KM, Sheppard A, Gluckman PD, *et al.* Epigenetic Gene Promoter Methylation at Birth Is Associated With Child's Later Adiposity. *Diabetes* 2011; **60**: 1528–34.

372 Relton CL, Groom A, St Pourcain B, *et al.* DNA methylation patterns in cord blood DNA and body size in childhood. *PloS One* 2012; **7**: e31821.

373 Perng W, Mora-Plazas M, Marín C, Rozek LS, Baylin A, Villamor E. A Prospective Study of LINE-1DNA Methylation and Development of Adiposity in School-Age Children. *PloS One* 2013; **8**: e62587.

374 Milagro FI, Campión J, Cordero P, *et al.* A dual epigenomic approach for the search of obesity biomarkers: DNA methylation in relation to diet-induced weight loss. *FASEB J Off Publ Fed Am Soc Exp Biol* 2011; **25**: 1378–89.

375 Barres R, Kirchner H, Rasmussen M, *et al.* Weight loss after gastric bypass surgery in human obesity remodels promoter methylation. *Cell Rep* 2013; **3**: 1020–7.

376 Moleres A, Campión J, Milagro FI, *et al.* Differential DNA methylation patterns between high and low responders to a weight loss intervention in overweight or obese adolescents: the EVASYON study. *FASEB J Off Publ Fed Am Soc Exp Biol* 2013. doi:10.1096/fj.12-215566.

377 Nichols H, Sandler D, Deroo L, Xu Z, Taylor J. Hormonal Risk Factors for Breast Cancer and DNA Methylation. *Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol* 2013; **22**: 471–2.

378 Riboli E, Hunt KJ, Slimani N, *et al.* European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutr* 2002; **5**: 1113–24.

379 Karimi M, Johansson S, Stach D, *et al.* LUMA (LUminometric Methylation Assay)--a high throughput method to the analysis of genomic DNA methylation. *Exp Cell Res* 2006; **312**: 1989–95.

380 Karimi M, Johansson S, Ekström TJ. Using LUMA: a Luminometric-based assay for global DNA-methylation. *Epigenetics Off J DNA Methylation Soc* 2006; **1**: 45–8.

381 Bjornsson HT, Sigurdsson MI, Fallin MD, *et al.* Intra-individual change over time in DNA methylation with familial clustering. *JAMA J Am Med Assoc* 2008; **299**: 2877–83.

382 Bibikova M, Barnes B, Tsan C, *et al.* High density DNA methylation array with single CpG site resolution. *Genomics* 2011; **98**: 288–95.

383 Eckhardt F, Lewin J, Cortese R, *et al.* DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat Genet* 2006; **38**: 1378–85.

384 Triche TJ, Weisenberger DJ, Berg DVD, Laird PW, Siegmund KD. Low-level processing of Illumina Infinium DNA Methylation BeadArrays. *Nucleic Acids Res* 2013. doi:10.1093/nar/gkt090.

385 Bolstad BM, Irizarry RA, Åstrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 2003; **19**: 185–93.

386 Dedeurwaerder S, Defrance M, Calonne E, Denis H, Sotiriou C, Fuks F. Evaluation of the Infinium Methylation 450K technology. *Epigenomics* 2011; **3**: 771–84.

387 Du P, Zhang X, Huang C-C, *et al.* Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* 2010; **11**: 587.

388 Maksimovic J, Gordon L, Oshlack A. SWAN: Subset-quantile Within Array Normalization for Illumina Infinium HumanMethylation450 BeadChips. *Genome Biol* 2012; **13**: R44.

389 Teschendorff AE, Marabita F, Lechner M, *et al.* A Beta-Mixture Quantile Normalisation method for correcting probe design bias in Illumina Infinium 450k DNA methylation data. *Bioinformatics* 2012. doi:10.1093/bioinformatics/bts680.

390 Roessler J, Ammerpohl O, Gutwein J, *et al.* Quantitative cross-validation and content analysis of the 450k DNA methylation array from Illumina, Inc. *BMC Res Notes* 2012; **5**: 210.

391 Touleimat N, Tost J. Complete pipeline for Infinium(®) Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. *Epigenomics* 2012; **4**: 325–41.

392 Chen C, Grennan K, Badner J, *et al.* Removing Batch Effects in Analysis of Expression Microarray Data: An Evaluation of Six Batch Adjustment Methods. *PLoS ONE* 2011; **6**: e17238.

393 Leek JT, Scharpf RB, Bravo HC, *et al.* Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet* 2010; **11**: 733–9.

394 Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 2007; **8**: 118–27.

395 Leek JT, Storey JD. Capturing Heterogeneity in Gene Expression Studies by Surrogate Variable Analysis. *PLoS Genet* 2007; **3**: e161.

396 Teschendorff AE, Zhuang J, Widschwendter M. Independent surrogate variable analysis to deconvolve confounding factors in large-scale microarray profiling studies. *Bioinforma Oxf Engl* 2011; **27**: 1496–505.

397 Marabita F, Almgren M, Lindholm ME, *et al.* An evaluation of analysis pipelines for DNA methylation profiling using the Illumina HumanMethylation450 BeadChip platform. *Epigenetics Off J DNA Methylation Soc* 2013; **8**: 333–46.

398 Price ME, Cotton AM, Lam LL, *et al.* Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. *Epigenetics Chromatin* 2013; **6**: 4.

399 Zhang X, Mu W, Zhang W. On the Analysis of the Illumina 450k Array Data: Probes Ambiguously Mapped to the Human Genome. *Front Genet* 2012; **3**. doi:10.3389/fgene.2012.00073.

400 Adalsteinsson BT, Gudnason H, Aspelund T, *et al.* Heterogeneity in white blood cells has potential to confound DNA methylation measurements. *PloS One* 2012; **7**: e46705.

401 Reinius LE, Acevedo N, Joerink M, *et al.* Differential DNA Methylation in Purified Human Blood Cells: Implications for Cell Lineage and Studies on Disease Susceptibility. *PLoS ONE* 2012; **7**: e41361.

402 Sandoval J, Heyn H, Moran S, *et al.* Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics Off J DNA Methylation Soc* 2011; **6**: 692–702.

403 Mill J, Heijmans BT. From promises to practical strategies in epigenetic epidemiology. *Nat Rev Genet* 2013; **14**: 585–94.

404 Clifford H, Wessely F, Pendurthi S, Emes RD. Comparison of clustering methods for investigation of genome-wide methylation array data. *Front Genet* 2011; **2**: 88.

405 Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinforma Oxf Engl* 2010; **26**: 1572–3.

406 Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2008; **4**: 44–57.

407 Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 2009; **37**: 1–13.

408 Ziller MJ, Gu H, Müller F, *et al.* Charting a dynamic DNA methylation landscape of the human genome. *Nature* 2013; **advance online publication**. doi:10.1038/nature12433.

409 Hubbard SR, Till JH. Protein tyrosine kinase structure and function. *Annu Rev Biochem* 2000; **69**: 373–98.

410 Chen P-Y, Ganguly A, Rubbi L, *et al.* Intra-Uterine Calorie Restriction Affects Placental DNA Methylation and Gene Expression. *Physiol Genomics* 2013. doi:10.1152/physiolgenomics.00034.2013.

411 Xu X, Gammon MD, Jefferson E, *et al.* The influence of one-carbon metabolism on gene promoter methylation in a population-based breast cancer study. *Epigenetics Off J DNA Methylation Soc* 2011; **6**: 1276–83.

412 Pogribny IP, Prizimirska TV, Kulik GI, *et al.* Age-related effects of methionine-enriched diet on plasma homocysteine concentration and methylation of hepatic DNA in rats. *Ukr Biokhimicheskiĭ Zhurnal* 2005; **77**: 114–9.

413 Choi S-W, Friso S, Ghandour H, Bagley PJ, Selhub J, Mason JB. Vitamin B-12 Deficiency Induces Anomalies of Base Substitution and Methylation in the DNA of Rat Colonic Epithelium. *J Nutr* 2004; **134**: 750 –755.

414 Bönsch D, Lenz B, Fiszer R, Frieling H, Kornhuber J, Bleich S. Lowered DNA methyltransferase (DNMT-3b) mRNA expression is associated with genomic DNA hypermethylation in patients with chronic alcoholism. *J Neural Transm Vienna Austria 1996* 2006; **113**: 1299–304.

415 ASIC1 Gene - GeneCards | ASIC1 Protein | ASIC1 Antibody. http://www.genecards.org/cgi-bin/carddisp.pl?gene=ASIC1 (accessed 15 Jul2013).

416 Elks CE, Perry JRB, Sulem P, *et al.* Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies. *Nat Genet* 2010; **42**: 1077–85.

417 Peng S, Lü B, Ruan W, Zhu Y, Sheng H, Lai M. Genetic polymorphisms and breast cancer risk: evidence from meta-analyses, pooled analyses, and genome-wide association studies. *Breast Cancer Res Treat* 2011; **127**: 309–24.

418 Choi J-Y, James SR, Link PA, *et al.* Association between global DNA hypomethylation in leukocytes and risk of breast cancer. *Carcinogenesis* 2009; **30**: 1889 –1897.

419 Xu X, Gammon MD, Hernandez-Vargas H, *et al.* DNA methylation in peripheral blood measured by LUMA is associated with breast cancer in a population-based study. *FASEB J* 2012; **26**: 2657–66.

420 Christensen BC, Kelsey KT, Zheng S, *et al.* Breast Cancer DNA Methylation Profiles Are Associated with Tumor Size and Alcohol and Folate Intake. *PLoS Genet* 2010; **6**: e1001043.

421 Flanagan JM, Cocciardi S, Waddell N, *et al.* DNA methylome of familial breast cancer identifies distinct profiles defined by mutation status. *Am J Hum Genet* 2010; **86**: 420–33.

422 Starlard-Davenport A, Tryndyak VP, James SR, *et al.* Mechanisms of epigenetic silencing of the Rassf1a gene during estrogen-induced breast carcinogenesis in ACI rats. *Carcinogenesis* 2010; **31**: 376–81.

423 Tanner JM. Trend towards Earlier Menarche in London, Oslo, Copenhagen, the Netherlands and Hungary. *Publ Online 11 May 1973 Doi101038243095a0* 1973; **243**: 95–6.

424 Mervish NA, Gardiner EW, Galvez MP, *et al.* Dietary flavonol intake is associated with age of puberty in a longitudinal cohort of girls. *Nutr Res N Y N* 2013; **33**: 534–42.

425 KLF9 Gene - GeneCards | KLF9 Protein | KLF9 Antibody. http://www.genecards.org/cgi-bin/carddisp.pl?gene=KLF9 (accessed 15 Jul2013).

426 Bock C. Analysing and interpreting DNA methylation data. *Nat Rev Genet* 2012; **13**: 705–19.

427 Chadeau-Hyam M, Ebbels TMD, Brown IJ, *et al.* Metabolic Profiling And The Metabolome-Wide Association Study: Significance Level For Biomarker Identification. *J Proteome Res* 2010; **9**: 4620–7.

428 Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity* 2005; **95**: 221–7.

429 Wild CP. The exposome: from concept to utility. *Int J Epidemiol* 2012. doi:10.1093/ije/dyr236.

430 Reik W, Dean W, Walter J. Epigenetic reprogramming in mammalian development. *Science* 2001; **293**: 1089–93.

431 Brennan K, Flanagan JM. Is there a link between genome-wide hypomethylation in blood and cancer risk? *Cancer Prev Res Phila Pa* 2012; **5**: 1345–57.

432 Wu H-C, Delgado-Cruzata L, Flom JD, *et al.* Global methylation profiles in DNA from different blood cell types. *Epigenetics Off J DNA Methylation Soc* 2011; **6**: 76–85.

433 Houseman EA, Accomando WP, Koestler DC, *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* 2012; **13**: 86.

434 Scoccianti C, Ricceri F, Ferrari P, *et al.* Methylation patterns in sentinel genes in peripheral blood cells of heavy smokers: Influence of cruciferous vegetables in an intervention study. *Epigenetics* 2011; **6**: 1114–9.

435 Relton CL, Smith GD. Is epidemiology ready for epigenetics? *Int J Epidemiol* 2012; **41**: 5–9.

436 Ng JW, Barrett LM, Wong A, Kuh D, Smith GD, Relton CL. The role of longitudinal cohort studies in epigenetic epidemiology: challenges and opportunities. *Genome Biol* 2012; **13**: 246.

437 Gordon L, Joo JE, Powell JE, *et al.* Neonatal DNA methylation profile in human twins is specified by a complex interplay between intrauterine environmental and genetic factors, subject to tissue-specific influence. *Genome Res* 2012; **22**: 1395–406.

438 Jaffe AE, Murakami P, Lee H, *et al.* Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. *Int J Epidemiol* 2012; **41**: 200–9.

439 Relton CL, Smith GD. Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. *Int J Epidemiol* 2012; **41**: 161–76.

440 Kirkbride JB, Susser E, Kundakovic M, Kresovich JK, Davey Smith G, Relton CL. Prenatal nutrition, epigenetics and schizophrenia risk: can we test causal effects? *Epigenomics* 2012; **4**: 303–15.

441 Vineis P, Perera F. Molecular epidemiology and biomarkers in etiologic cancer research: the new in light of the old. *Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol* 2007; **16**: 1954–65.

APPENDIX A

The mediterranean dietary pattern and breast cancer risk in Greek-Cypriot women: a case-control study

BMC Cancer

RESEARCH ARTICLE Open Access

The mediterranean dietary pattern and breast cancer risk in Greek-Cypriot women: a case-control study

Christiana A Demetriou[1,2*], Andreas Hadjisavvas[1], Maria A Loizidou[1], Giorgos Loucaides[3], Ioanna Neophytou[1], Sabina Sieri[4], Eleni Kakouri[5], Nicos Middleton[6], Paolo Vineis[2] and Kyriacos Kyriacou[1]

Abstract

Background: Diet has long been suspected to impact on breast cancer risk. In this study we evaluated whether the degree of adherence to a Mediterranean diet pattern modifies breast cancer risk amongst Greek-Cypriot women.

Methods: Subjects included 935 cases and 817 controls, all participating in the MASTOS case-control study in Cyprus. The study was approved by the Cyprus National Bioethics Committee. Information on dietary intakes was collected using an interviewer administered 32-item Food Frequency Questionnaire. Information on demographic, anthropometric, lifestyle, and other confounding factors was also collected. Adherence to the Mediterranean Diet pattern was assessed using two a-priory defined diet scores. In addition, dietary patterns specific to our population were derived using Principal Component Analysis (PCA). Logistic regression models were used to assess the association between the dietary patters and breast cancer risk.

Results: There was no association with breast cancer risk for either score, however, higher consumptions of vegetables, fish and olive oil, were independently associated with decreased risk. In addition, the PCA derived component which included vegetables, fruit, fish and legumes was shown to significantly reduce risk of breast cancer (ORs across quartiles of increasing levels of consumption: 0.89 95%CI: 0.65-1.22, 0.64 95%CI: 0.47-0.88, 0.67 95%CI: 0.49-0.92, P trend < 0.0001), even after adjustment for relevant confounders.

Conclusions: Our results suggest that adherence to a diet pattern rich in vegetables, fish, legumes and olive oil may favorably influence the risk of breast cancer. This study is the first investigation of dietary effects on breast cancer risk in Cyprus, a country whose population has traditionally adhered to the Mediterranean diet.

Background

Breast Cancer (BC) is the most prevalent cancer among females worldwide, being responsible for more than 515 thousand deaths annually in all WHO regions [1]. Diet has long been suspected to affect BC risk and numerous studies have investigated its possible effect. Even though most of these studies conclude that any effect of diet on risk may be weak, the findings of some large scale prospective studies [2-10] emphasize the importance of further investigating the role of diet in BC etiology.

When investigating the effects of diet on BC, it is important to consider that individuals do not consume single foods, but combinations of several foods that contain both nutrient and non-nutrient substances. Given the complexity of human diets, the correlation and effect modification of intake of some nutrients, and the many nutrient-to-nutrient interactions, conclusions about the effect of consumption of a single nutrient, food group, or dietary constituent on a specific health outcome may be misleading. For these reasons, it is useful to examine patterns of nutrient intake that express several related aspects of dietary intake concurrently [11,12].

* Correspondence: christianad@cing.ac.cy
[1]Department of EM/Molecular Pathology, The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus
Full list of author information is available at the end of the article

BioMed Central

© 2012 Demetriou et al; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

One such dietary pattern is the Mediterranean diet pattern. The Mediterranean diet has long been accepted as an example of a well-balanced diet but there is no gold standard "Mediterranean diet". At least 16 countries border the Mediterranean Sea and diets vary between these countries as well as also between regions within the same country. However, despite the above variation, a Mediterranean diet has the following common characteristics: high consumption of fruits, vegetables, bread and other cereals, potatoes, beans, nuts and seeds; low to moderate consumption of dairy products, fish, eggs and poultry; and infrequent consumption of red meat. In the Mediterranean diet, olive oil is frequently consumed and is an important mono-unsaturated fat source, and wine is consumed in low to moderate amounts.

The combination and range of foods included in this diet pattern, provide plenty of anti-oxidants such as flavonoids, carotenoids, and antioxidant vitamins, lots of phytochemicals including phytoestrogens, sufficient quantities of fiber, adequate folate, and a favourable fatty acid profile [10,13]. These nutrients have been linked to mechanisms of carcinogenesis and have been found [14], or are hypothesized to confer protective effects [15] on total cancer incidence, and more specifically on BC incidence. The hypothesis that adherence to a Mediterranean diet pattern decreases BC risk is also supported by the observation that countries bordering the Mediterranean Sea, which are more likely to adhere to such a diet, for example Greece, Spain, Italy and Cyprus, have the lowest BC incidence rates in Europe [1].

Adherence to the Mediterranean diet pattern in epidemiological studies is commonly assessed with the use of dietary scores, developed to capture how closely a subject's diet resembles the Mediterranean diet. In this study we evaluated whether the degree of adherence to a Mediterranean diet pattern is associated with a decreased BC risk amongst Greek-Cypriot women who participated in the MASTOS (Greek for "breast") study - to date, the largest case-control study carried out in Cyprus [16]. While it is commonly assumed that traditionally a large proportion of the island's population adheres to a Mediterranean diet pattern, the extent to which this is true is unclear, let alone whether, and which components, may be associated with a reduced BC risk, making first-time investigation of this issue among the Mediterranean island's female population invaluable.

Methods

Subjects

MASTOS was the first and to date largest BC case-control study to be carried out in Cyprus. Recruitment lasted for the three year period 2004-2006, and the main aim of the MASTOS study was to describe the frequency of established and recognized risk factors for BC among Cypriot women and to assess their association with BC occurrence. Information on the study design, data collection as well as initial analysis of MASTOS data with respect to BC risk factors is published elsewhere [16]. In summary, 1109 cases of breast cancer aged 40-70 years old and 1177 control women were recruited. Cases had a confirmed diagnosis of breast cancer between January 1999 and December 2005. Controls were women from the general population that attended the national mammography screening program and received a negative result. In the present report, the analysis was restricted to post-menopausal cases and controls only, since diet has been shown to have a different effect on pre- and post-menopausal BC [10,17,18].

Data collection

Demographic and risk factor data were collected with the use of a specially designed questionnaire, through a standardized interview. The questionnaire included information on age, marital status, level of education, current and adolescent BMI, exercise status, smoking status, family history of breast or ovarian cancer, hormone use, and reproductive characteristics such as age at menarche, pregnancy, gestation period, parity, breastfeeding, age at first and last pregnancy. Weight, height and consequently BMI were based on actual measurements but all other information was self-reported. Exercise was taken to mean any form of regular recreational exercise, 2-3 times a week, for a minimum of 30 minutes each time.

In addition, a standardized diet interview was conducted with each subject using a Food Frequency Questionnaire (FFQ) designed to best capture consumption of food items whose contribution to the energy and nutrient intake of the Cypriot population is high. The resulting questionnaire was designed to capture an individual's habitual food intake over the previous year. Thirty two food items were included in the questionnaire under eight distinct categories i.e. fruit and vegetables, fish, meat, meat products and substitutes, cereals, dairies, agricultural products and sweets. Frequency of consumption for each food item was reported as number of portions daily, weekly or monthly, rarely consumed, or never consumed. Portion sizes for the foods consumed were assessed by showing subjects pictures of standard medium portions, decided on by experienced dieticians. The questionnaire also included questions on the kind of oil used while cooking, whether olive oil was used in salads, and what kinds of spreads were used on bread.

Adherence to the Mediterranean diet pattern was first assessed using the diet score proposed by Panagiotakos et al. [19]. The score was developed to incorporate the inherent characteristics of the Mediterranean dietary pattern and to examine its relationship with cardiovascular disease risk and biological markers [19]. Thus, the score includes the weekly consumption of the following 9 food groups: non-refined cereals, potatoes, fruit, vegetables, legumes, fish, red meat and products, poultry, full fat dairy products as well as use of olive oil in cooking, and alcoholic beverages. The scoring system for the level of consumption of each food group is described in Table 1 and the final score ranged from 0 to 55. Caloric intakes were not available due to the short nature of the FFQ and the grouping of several different foods in the same food group category. As a result, BMI and height were used as surrogate measures of the energy needs of each individual and were thus included and corrected for in all analyses. For the same reason, nutrient intakes could not be estimated with sufficient accuracy; therefore the diet scores chosen were based on food group intakes and did not involve any nutrient components.

Some of the original diet score categories were modified for our analysis. First, the Cereals category was taken to include all cereals consumed, as information on types of cereal was not available. Unrefined and refined cereals differ mostly in their fiber content [20], therefore since nutrient measures were not estimated it was thought appropriate to group together all cereal consumption. In addition, use of olive oil in cooking was estimated to include use in cooking and as a salad dressing, as usage of olive oil in salads is the most common use of olive oil in Cyprus. Lastly, the Mediterranean Diet Score was first designed to assess the effect of diet on cardiovascular disease. Since some alcohol consumption is beneficial to cardiovascular health, in the original score, zero alcohol consumption received a low score. However, for cancer, any amount of alcohol is detrimental to cancer risk [8]. Therefore, for our purposes zero alcohol consumption, received the highest score.

Secondly, the diet score developed by Martinez-Gonzalez et al. [21] was used. The score was developed as a quicker and simpler tool to estimate quantitatively adherence to the cardioprotective elements of the Mediterranean diet [21]. Nine items are included in the diet score and scored as shown in Table 2. The resulting score ranged from 0 to 9 points.

Again, some of the original score components were modified. Wine consumption (component 7) was substituted by any alcohol consumed. Furthermore, since this score was also originally designed to assess diet's effects on cardiovascular health, the scoring was reversed to score +1 if no alcohol was consumed. Lastly, one of the characteristics of the Mediterranean diet is high consumption of cereals including pasta and potatoes. Therefore, component nine included pasta, rice and potatoes.

Table 1 The Mediterranean Diet Score by Panagiotakos et al. as modified for the present analysis

How often do you consume:	Frequency of consumption (servings/week or otherwise stated)					
All cereals (bread, pasta, rice, etc.)	Never	1-6	7-12	13-18	19-31	> 32
	0	1	2	3	4	5
Potatoes	Never	1-4	5-8	9-12	13-18	> 18
	0	1	2	3	4	5
Fruits	Never	1-4	5-8	9-15	16-21	> 22
	0	1	2	3	4	5
Vegetables	Never	1-6	7-12	13-20	21-32	> 33
	0	1	2	3	4	5
Legumes	Never	< 1	1-2	3-4	5-6	> 6
	0	1	2	3	4	5
Fish	Never	< 1	1-2	3-4	5-6	> 6
	0	1	2	3	4	5
Red Meat Products	≤ 1	2-3	4-5	6-7	9-10	> 10
	5	4	3	2	1	0
Poultry	≤ 3	4-5	5-6	7-8	9-10	> 10
	5	4	3	2	1	0
Full fat dairy products (cheese, yogurt, milk)	≤ 10	11-15	16-20	21-28	29-30	> 30
	5	4	3	2	1	0
Use of Olive Oil (times/week)	Never	Rare	< 1	1-3	3-5	Daily
	0	1	2	3	4	5
Alcohol (ml/day)	0	< 300	400	500	600	> 700
	5	4	3	2	1	0

Table 2 Dietary Items included in the Martinez-Gonzalez Score as modified for the present analysis

	Yes
1. Olive Oil (≥ 1 spoon/day)	+1
2. Fruit (≥ 1 serving/day)	+1
3. Vegetables *or* Salad (≥ 1 serving/day)	+1
4. Fruit (≥ 1 serving/day) *and* Vegetables (≥ 1 serving/day)[a]	+1
5. Legumes (≥ 2 servings/week)	+1
6. Fish (≥ 3 servings/week)	+1
7. Any alcohol (zero consumption)	+1
8. Meat (< 1 serving/day)	+1
9. All cereals (bread, rice etc.) (1-3 servings/day)	+1

[a] One point is added when ≥ serving/day of both fruits and vegetables is consumed

For the purposes of this analysis consumption of one to three portions of any cereal daily was scored with +1 and all other consumption (i.e. lower or higher) was scored with zero.

Statistical analysis

Differences between socio-demographic characteristics, potential risk factors (including overall diet scores) between cases and controls were initially explored using chi-squared or *t*-test as appropriate (e.g. height). In order to assess the association between levels of adherence to a Mediterranean diet and BC, adjusted Odds Ratios (ORs) of BC risk were estimated (a) across categorical levels and (b) per 1 unit increase in either of the diet scores in multivariable logistic models after adjusting for the possible confounding effect of the risk factors that best predict BC risk in this population [16]. In addition to the overall scores, ORs of BC per 1 unit increase in servings/week (or servings/day accordingly) were also estimated for each single components of the diet scores.

To further examine the dietary intakes of the Cypriot population, factor analysis with principal component extraction (PCA) was performed in order to derive the patterns that best describe the dietary habits of our population. It was deemed appropriate to restrict this analysis to the control group. Dietary intakes (in g/month for food items and ml/month for drink items) of each of the 32 food and beverage items in our FFQ were considered. Olive oil was not entered as a variable in PCA because only frequency and not quantity of consumption was assessed in the FFQ. Extraction of principal components was followed by a diagonal (oblimin) rotation, following failure of orthogonal rotations to produce interpretable results. For each dietary pattern identified, a component score was calculated for each subject using all factor loadings and the respective consumption of each food and beverage item. The association between BC and each of the observed factors were investigated in multivariable logistic models after mutually adjusting for all others. All statistical analyses were performed using STATA statistical software, version 11 (StatCorp. 2007. College Station, TX) and SPSS statistical software version 17.0 (PASW Inc.)

Results

The characteristics of the 1752 post-menopausal women (among a total of 2286, 76.6%) participating in the MASTOS case-control study are presented in Table 3. As expected, compared to controls, cases tend to be more educated, of a larger BMI, an older age at FFTP, more likely to have a family history of BC, less likely to have breastfed their children, more likely to have smoked and less likely to have exercised more than 6 hours per week prior to diagnosis. However, contrary to what was expected, cases were younger and less likely to have used HRT. Cases and controls did not differ in terms of height, age at menarche, parity. Furthermore, no association was observed between case/control status and either of the two diet scores investigated here.

However, the distribution of cases and controls across the diet score categories indicated that the Mediterranean diet pattern is adhered to by a large percentage of Cypriots. More than 40% of the Cypriot population has a score of above 35 using the Panagiotakos Mediterranean diet score, and close to 40% of the Cypriot population scores above 6 using the Martinez-Gonzalez score.

Mediterranean diet score by Panagiotakos

The mean intake of the 11 food indices used to construct the Mediterranean Diet score by Panagiotakos among post-menopausal cases and controls is shown in Table 4. Risk of BC decreased in association with high intake of vegetables and salads (OR per unit increase in servings/week: 0.95, 95% CI: 0.92-0.99), fish (OR per unit increase in servings/week: 0.88, 95% CI: 0.79-0.98), and olive oil (OR per unit increase in servings/week: 0.95, 95% CI: 0.92-0.99). A unit increase in the overall diet score by Panagiotakos was not associated with risk of post-menopausal BC (Table 4). Similarly, when treated as a categorical variable, divided in quartiles, this diet score was not associated with BC risk (OR Q4 vs. Q1: 0.99 95% CI: 0.70-1.40) (Table 4).

Mediterranean diet score by Martinez-Gonzalez

The mean intake of the 9 food indexes used to construct the Mediterranean Diet score by Martinez Gonzalez among post-menopausal cases and controls is shown in Table 5. Similar to what was observed with the diet score by Panagiotakos, risk of BC decreased in association with high intake of vegetables and salads (OR per unit increase in servings/day: 0.74, 95% CI: 0.57-0.96), fish (OR per unit increase in servings/week: 0.88, 95% CI: 0.79-0.98), and olive oil (OR per unit increase in

Table 3 Sociodemographic characteristics and potential risk factors among post-menopausal participants of the MASTOS case control study

Variable	Controls (% of n = 817)[a]	Cases (% of n = 935)[a]	P-Value
Education Category			
- Primary	395 (48.3)	394 (42.1)	0.044
- Secondary	277 (33.9)	334 (35.7)	
- College	77 (9.4)	98 (10.5)	
- University	62 (7.6)	96 (10.3)	
Age at Interview			
- < 45 years	4 (0.5)	57 (6.1)	< 0.001
- 45-54 years	187 (22.9)	263 (28.1)	
- 55-64 years	488 (59.7)	383 (41.0)	
- ≥ 65	138 (16.9)	232 (24.8)	
Height	158.7 ± 6.2	159.1 ± 7.1	0.190[b]
BMI category			
- < 20 (kg/m^2)	6 (0.7)	20 (21.4)	0.001
- 20-24.9 (kg/m^2)	190 (23.3)	216 (23.1)	
- 25-29.9 (kg/m^2)	323 (39.5)	326 (34.9)	
- ≥ 30 (kg/m^2)	201 (24.6)	303 (32.4)	
Family History of BC			
- No	750 (91.8)	770 (82.4)	< 0.001
- Yes	66 (8.1)	163 (17.4)	
Age at Menarche			
- ≥ 14	326 (39.9)	331 (35.4)	0.087
- 12-13	400 (49.0)	477 (51.0)	
- > 12	87 (10.6)	123 (13.2)	
Parity			
- 0	65 (8.0)	88 (9.4)	0.251
- 1	56 (6.9)	68 (7.3)	
- 2	330 (40.4)	403 (43.1)	
- 3 or more	366 (44.8)	376 (40.2)	
Age at FFTP[c]			
- < 20	108 (13.2)	92 (9.8)	0.007
- 20 - < 25	378 (46.3)	397 (42.5)	
- 25 - < 30	189 (23.1)	236 (25.2)	
- ≥ 30	123 (15.1)	185 (19.8)	
Breastfeeding			
- Yes	574 (70.3)	606 (64.8)	0.015
- No	243 (29.7)	329 (35.2)	
Oral Contraceptive use			
- No	679 (83.1)	758 (81.1)	0.352
- Yes	137 (16.8)	172 (18.4)	
HRT use			
- No	572 (70.0)	794 (84.9)	< 0.001
- Yes	238 (29.1)	133 (14.2)	
Smoking Status			
- Never	676 (82.7)	747 (79.9)	0.003
- Past	53 (6.5)	102 (10.9)	
- Current	88 (10.8)	84 (9.0)	
Exercise			
- No exercise	399 (48.8)	523 (55.9)	< 0.001
- Up to 3 hrs/wk	214 (26.2)	252 (27.0)	

Table 3 Sociodemographic characteristics and potential risk factors among post-menopausal participants of the MASTOS case control study *(Continued)*

- > 3 and ≤ 6 hrs/wk	141 (17.3)	127 (13.6)	
- > 6 hrs/wk	63 (7.7)	33 (3.5)	
Mediterranean Diet Score - Panagiotakos			
- ≤ 30	227 (27.8)	292 (31.2)	0.264
- 31-35	224 (27.4)	266 (28.4)	
- 36-40	239 (29.3)	246 (26.3)	
- > 40	127 (15.5)	131 (14.0)	
Mediterranean Diet Score - Martinez-Gonzalez			
- 0-4	258 (31.6)	322 (34.4)	0.429
- 5-6	237 (29.0)	242 (25.9)	
- 6-7	196 (24.0)	221 (23.6)	
- 8-9	126 (15.4)	150 (16.1)	

[a] Failure of category counts to add up to this number, or of percentages to add up to 100 denotes missing values

[b] Mean (SD) and p-value of independent t-test. In all other cases, p-values of χ^2 test are presented.

[c] First full-term pregnancy

spoons/day: 0.76, 95% CI: 0.59-0.97). However, a unit increase in the diet score by Martinez-Gonzalez did not significantly increase or decrease risk of post-menopausal BC (Table 5). Similarly, when treated as a categorical variable, in quartiles, this diet score was not associated with BC risk (OR Q4 vs. Q1: 0.99 (0.70-1.40).

Principal component analysis

Based on an eigenvalue criterion of > 1.0, and scree plot analysis 11 components were initially observed. However, factor interpretability justified the retention of only 4 components, as components 5 to 11 showed high factor loadings on single variables. As a result we repeated

Table 4 Mediterranean Diet Score by Panagiotakos et al. and Post-Menopausal BC risk

	Intake[a, b]		
Food Indexes	**Post-menopausal cases (n = 934)**	**Post-menopausal controls (n = 815)**	**Adjusted OR per unit increase (95% CI)[c]**
All cereals	18.7 ± 7.8	19.4 ± 8.8	0.99 (0.98, 1.00)
Potatoes	2.0 ± 1.4	1.9 ± 1.3	0.95 (0.87, 1.03)
Fruits	9.2 ± 6.0	9.6 ± 6.9	1.00 (0.99, 1.02)
Vegetables	5.8 ± 2.8	6.2 ± 3.4	0.95 (0.92, 0.99)*
Legumes	2.1 ± 1.2	2.2 ± 1.4	0.95 (0.88, 1.03)
Fish	1.6 ± 1.1	1.8 ± 1.2	0.88 (0.79, 0.98)*
Red meat and products	2.8 ± 2.1	2.7 ± 2.1	0.98 (0.93, 1.03)
Poultry	1.6 ± 0.9	1.7 ± 0.9	0.91 (0.81, 1.04)
Full fat dairy products	6.5 ± 6.2	7.3 ± 6.8	0.98 (0.96, 1.00)
Use of olive oil (times/week)	6.6 ± 2.9	7.0 ± 3.4	0.95 (0.92, 0.99)*
Alcohol	16.7 ± 43.1	18.1 ± 49.5	1.00 (1.00, 1.00)
Mediterranean Diet Score	35.5 ± 3.1	35.8 ± 2.9	1.00 (0.96, 1.04)
Mediterranean Diet Score Across quartile levels			
- Q1: ≤ 34			1.00
- Q2: 35-36			1.02 (0.76, 1.36)
- Q3: 37-38			1.00 (0.75, 1.34)
- Q4: > 38			0.99 (0.70, 1.40)
			p-value for trend = 0.06

Levels (means ± SDs) of the Mediterranean Diet Score food indices in post-menopausal breast cancer cases and post-menopausal controls and adjusted odds ratios (with 95% CI) of BC in terms of the overall score and each its components

[a] All values are means ± SDs

[b] Servings/week or as otherwise stated

[c] Adjusted for the following covariates: age at interview, family history, age at FFTP, HRT use, exercise, age at menarche, height and BMI in post-menopausal women

Table 5 The Martinez-Gonzalez Diet Score and Post-Menopausal BC risk

	Intake[a]		
Food Indexes	**Post-menopausal cases (n = 934)**	**Post-menopausal controls (n = 815)**	**Adjusted OR per unit increase (95% CI)[b]**
Olive oil (servings/day)	1.1 ± 0.4	1.1 ± 0.5	0.76 (0.59, 0.97)*
Fruit (servings/day)	1.3 ± 0.8	1.3 ± 0.9	1.01 (0.89, 1.15)
Vegetables or Salad (servings/day)	0.8 ± 0.4	0.9 ± 0.5	0.74 (0.57, 0.96)*
Fruit AND Vegetables (servings/day)	2.1 ± 1.0	2.2 ± 1.1	0.96 (0.86, 1.06)
Legumes (servings/week)	2.1 ± 1.2	2.2 ± 1.4	0.94 (0.87, 1.02)
Fish (servings/week)	1.6 ± 1.1	1.8 ± 1.2	0.88 (0.79, 0.98)*
Meat (servings/day)	0.6 ± 0.3	0.6 ± 0.3	0.73 (0.54, 1.04)
Cereals (servings/day)	2.9 ± 1.1	3.0 ± 1.2	0.93 (0.85, 1.02)
Alcohol (glasses/day)	0.1 ± 0.3	1.1 ± 0.4	0.99 (0.74, 1.31)
Martinez-Gonzalez Diet Score	5.5 ± 1.9	5.6 ± 1.9	1.00 (0.95, 1.06)
Martinez Gonzalez Diet Score			
Across quartile levels			
- Q1: 0-4			1.00
- Q2: 5-6			0.25 (0.64, 1.13)
- Q3: 6-7			0.84(0.72, 1.30)
- Q4: 8-9			0.63 (0.77, 1.53)
			p-value for trend = 0.61

Levels of the Martinez-Gonzalez Diet Score food indices in post-menopausal breast cancer cases and post-menopausal controls and adjusted odds ratios of BC (with 95% CI) in terms of the overall score and each of its components

[a]All values are means ± SDs

[b] Adjusted for the following covariates: age at interview, family history, age at FFTP, HRT use, exercise, age at menarche, height and BMI in post-menopausal women only.

the technique, retaining and rotating only four factors. Each of the retained components was labeled according to the dietary pattern they represented: Meat/Potatoes (pattern 1), Cereals/Milk/Dairy (pattern 2), Cakes/Sweets/Nuts/Crackers/Pasta/Rice (pattern 3) and Fruit/Vegetables/Fish/Legumes (pattern 4) (Table 6). The four factors accounted for 23.6% of the total variance in the 32 food and beverage items (8.05, 5.92, 5.10, 4.55% for the 4 patterns respectively).

Pattern 4 more closely resembles the Mediterranean diet pattern. This was further analyzed to determine its association to BC risk. Quartile values for Pattern 4 based on the distribution of controls, were used to divide cases and controls with the highest score quartile assigned a score of 4, while the lowest score quartile were assigned a score of 1. Quartiles of adherence to this dietary pattern were entered into a multivariate regression model adjusting for all potential confounders, to determine whether high adherence to the Fruit/Vegetables/Fish/Legumes dietary pattern decreases BC risk. Since, in the oblique rotation, the factors are allowed to be correlated, in the multivariable regression model we also adjusted for the other three PCA derived patterns, also as categorical variables. It was found that post-menopausal BC risk decreased across quartile levels of increased adherence to the fourth dietary pattern (OR Q4 vs. Q1: 0.67 95% CI: 0.49-0.92, P for trend < 0.0001) (Table 7). Patterns 1, 2, and 3 were not associated with BC risk whether treated as continuous or categorical variables, and there was no evident trend (data not shown).

Discussion

This study was the first to investigate dietary patterns in relation to BC risk in Cyprus, an island where the diet closely resembles the Mediterranean diet, as shown by the relatively large percentage of Cypriots that receive high scores of adherence to the Mediterranean diet patterns examined. None of the two Mediterranean Diet scores employed in our analysis were found to be associated with BC risk. However, in both scores, specific food indexes - Vegetables, Fish and Olive Oil - were inversely associated with BC risk. Although the individual effect estimates observed are small, if one assumes that the positive effects of different food groups are additive, then the reduction in BC risk through a diet that includes all the beneficial components, could be substantial. This combined effect becomes obvious when the fourth PCA dietary pattern, a combination of the protective components, is analyzed. Further, considering

Table 6 Principal Component Analysis Dietary Patterns

Food	Pattern 1: Meat/ Potatoes	Pattern 2: Cereals/Milk/ Dairy	Pattern 3: Cakes/Sweets/Nuts/Crackers/ Pasta/Rice	Pattern 4: Fruit/Vegetables/Fish/ Legumes
Red Meat	0.568			
Poultry	0.447			
Deli	0.445			
Potatoes	0.414			
Sausages	0.355			
Rabbit	0.342			
Yogurt Light		0.593		
Cheese Diet		0.546		
Milk Skimmed		0.430		
Cheese Low-fat		0.388		
Milk Full-fat		0.345		
Cheese Full-fat		0.342		
Cereals		0.342		
Cakes			0.621	
Chocolates			0.542	
Bakery			0.450	
Nuts			0.368	
Crackers/ Biscuits			0.362	
Pasta/Rice			0.345	
Fish fillet				0.565
Vegetables/ Salad				0.528
Fruit				0.447
Legumes				0.415
Fish other				0.336

Factor loadings > 0.340 based on Principal Components Analysis of average intake of 32 food and beverage items, standardized and followed by diagonal rotation

the high incidence of BC and assuming a causative association, even a small reduction in risk conferred by dietary components can be translated into several prevented cases per year, in Cyprus alone.

The application of Mediterranean diet scores in epidemiological studies has yielded conflicting results. Recently a study in the EPIC cohort, demonstrated that higher adherence to a Mediterranean dietary pattern - as assessed using the Mediterranean diet score developed by Trichopoulou [22] - is associated with a decreased cancer risk both in Mediterranean and in non-Mediterranean countries (HR: 0.96, 95% CI 0.96-0.98 for a two point increment in the score). This analysis did not identify any of the components of the score as having a predominant effect, supporting the idea that it is the combination and range of nutrient and non nutrient components of the diet, that confer the protective effect [13]. In a more focused analysis of the EPIC cohort with respect to BC risk, adherence to the Mediterranean diet pattern - assessed using an adaptation of the Trichopoulou score - was found to confer a small decrease in total and post-menopausal BC risk. This protective effect became stronger when no consumption of alcohol - instead of moderate -was treated as most beneficial and was more pronounced for ER-/PR- post-menopausal tumours [10]. A study on the Greek EPIC cohort has also demonstrated a marginally significant inverse association between adherence to a Mediterranean Diet pattern (defined a priori) and post-menopausal BC [23]. However, Cade et al. [24] who examined the adherence to a Mediterranean diet pattern in a British cohort did not yield any significant results, except a non significant trend towards decreased risk, with increasing adherence in pre-menopausal women. This cohort included women with different dietary patterns and at extremes of intake, so as to enable comparisons

Table 7 Odds Ratios of BC in relation to Principal Component Analysis-derived pattern 4

	Cases/ Controls	Adjusted OR[a]	95% Confidence Interval
Pattern 4: Fruit/Vegetables/Fish			
Quartile 1	261/174	1.00	
Quartile 2	251/192	0.89	0.65-1.22
Quartile 3	213/222	0.64	0.47-0.88
Quartile 4	208/226	0.67	0.49-0.92
P for trend		< 0.0001	

Quartiles of adherence to dietary pattern 4, derived from Principle Component Analysis, in relation to BC risk

[a] Odds Ratio adjusted for the following covariates: age at interview, family history, age at FFTP, HRT use, exercise, age at menarche, height, BMI and PCA derived patterns 1,2 and 3 in post-menopausal women only

and maximize the potential of successful identification of any interactions. Similarly, Fung et al. [2] investigated post-menopausal BC risk in the Nurses' Health Study, but found no significant association with adherence to the Mediterranean diet, however results for ER- BC approached significance. A meta-analysis of 6 prospective studies with a total of 512,366 subjects followed for a time ranging from eight to 18 years, showed that adherence to the Mediterranean diet is inversely associated with total cancer incidence and mortality, but there were no results specific to BC [25].

These conflicting findings might be explained by the limitations of diet scores such as the subjectivity associated with which food items are included in each score and how each component is scored. The score may be missing food items that largely determine the dietary habits of the population under study. Also, if the cut offs used for scoring are based on median values from a particular population, it is unlikely that they will yield useful and interpretable results, in a different population with different dietary habits. Another explanation for the contradictory findings is that traditional methods of food preparation vary widely within and between countries. It is argued that results from Northern European countries may differ from results from Mediterranean populations because essential components of the Mediterranean diet may be consumed in different forms and proportions in these two regions [24].

In order to derive dietary patterns that best reflect the dietary habits of the Cypriot population, PCA was applied to the dietary data, yielding four dietary patterns. The fourth dietary pattern - Fruit/Vegetables/Fish/Legumes - included both vegetables and fish that were found to be inversely associated with BC risk in both diet scores. In the MASTOS study, vegetable and fish intakes are correlated with olive oil consumption (0.9705 and 0.2180 respectively, both significant at the 0.05 level), therefore Pattern 4 can also be taken to include olive oil use despite the food's absence from the 32 FFQ variables entered into PCA. High adherence to this dietary pattern was found to decrease BC risk in post-menopausal women and can be recommended to women as a healthy dietary pattern which may indeed reduce BC risk.

It should be noted that PCA, is not without limitations since there is subjectivity associated with the derivation of the variables to be included, the number of factors extracted, the rotation used, and the labelling of the factors retained. Despite these, most studies applying PCA on their dietary data and deriving a dietary pattern similar to ours, demonstrated significant reductions in BC risk. While Velie et al. [12] found that a vegetable-fish/poultry-fruit diet pattern was not significantly correlated with BC, Ronco et al. [26] found that adherence to a food pattern rich in anti-oxidants - associated with white meat, fruit, vegetables - was protective in a study carried out in Uruguay. Murtaugh et al. [27] in their study of Hispanic and non-Hispanic white women, evidenced that a dietary pattern labeled as Mediterranean was statistically significant showing an inverse association with BC. In the French subgroup of the EPIC cohort, a "healthy/Mediterranean" (essentially vegetables, fruits, seafood, olive oil, and sunflower oil) diet pattern, was significantly inversely associated with BC risk, especially for ER+/PR- cancers [28]. Lastly, in the ORDET cohort, a dietary pattern which was characterized by high consumption of raw vegetables and olive oil as added fat, was found to significantly decrease BC risk [29].

The variables loading on the fourth PCA-derived pattern presented here have yielded varying results in studies investigating their relationship to BC. Concerning vegetable intake, most studies on post-menopausal women gave no significant findings [30-35]. However, Sonestedt et al. in an analysis of the Malmo Diet and Cancer Cohort demonstrated that in post-menopausal women with a BMI less than 27 kg/m^2 and in post-menopausal women not changing their diet before recruitments, there was a significant reduction in BC risk with increased intake of fruit, berries and vegetables [36]. Also, analysis of the Greek cohort of the EPIC study evidenced that a high consumption of vegetables and fruits was one of the most beneficial components, for the protective effect of the Mediterranean diet on survival [3]. In addition, in the EPIC study after a mean of 11 years of follow up, there was a significant reduction in BC incidence, associated with higher intake of vegetables [10]. Studies on fish intake again yielded mostly no significant results [37-43]. The only statistically significant result concerning fish intake comes from the Singapore Chinese Health Study, which evidenced that intake of marine n-3 fatty acids, had a

protective effect on BC incidence [44]. Lastly, concerning olive oil intake, a meta-analysis of 13800 patients and 23340 controls in 19 observational studies, evidenced that olive oil intake is inversely associated with BC risk [18]. In light of these findings, vegetables, fish, and olive oil may work synergistically in a dietary pattern to decrease BC risk, as was shown by our results.

Having discussed the significance of our findings in the context of the available literature, the limitations of this study should be mentioned. Cases included in the analysis were diagnosed over a seven year time frame, immediately before the conduction of the study. This may have introduced a survival bias into our study, but the 95% 10-year survival rate observed at the cancer referral centres in Cyprus [45] suggests that this bias is likely small. In addition, response rate was 98%, therefore selection bias other than survival is not a big concern. Further, our FFQ was limited since it examined intake of only 32 food items and there is no information about the way each food is consumed. Perhaps of most concern is the problem of measurement error in dietary assessment, particularly associated with the use of FFQs in case-control studies. However, it is unlikely that cases differentially recalled the specific food groups that were included in the diet scores or the dietary patterns derived from PCA. In addition, because diet is not an established BC risk factor, differential misclassification is less likely. On the other hand, both cases and controls are likely to have underreported their intakes, a common concern of FFQ use. However, such bias is non differential which tends to bias any findings towards the null, thus suggesting that the true effects might be stronger than those observed. Lastly, the four components retained from PCA only explained 23.6% of the variance, but this is comparable to other dietary studies, reflecting the challenges associated with reducing highly interrelated and complex dietary variables [46].

Despite the limitations associated with a case-control study design, these are the only data available in Cyprus since a cohort study has not been yet established in our population. Considering the location of the country and the dietary habits of the population that closely resemble the traditional Mediterranean diet, an investigation of this diet's effects on BC was warranted. Also, the sample size of this study ensures adequate power to delineate diet's effects on BC risk.

In addition to the association with BC risk, in our analysis there was a significant inverse association between adherence to the fourth PCA dietary pattern and age at interview (P value for chi-square test: 0.004). Thus, in the future, it will be interesting to investigate how the shift towards or away from this healthy dietary pattern in the younger generations impacts on BC incidence, as well as on other BC associated parameters.

Conclusion

In conclusion, our results suggest that adherence to a diet pattern rich in vegetables, fish, legumes and olive oil may favorably influence the risk of BC by having a protective effect. Despite the weak individual effect estimates, when these healthy food groups are combined into a dietary pattern the effect estimates obtained are larger and suggest that they act synergistically to confer a substantial reduction in BC risk. Given this protective effect in postmenopausal women, such a dietary pattern can be recommended to women as a healthy dietary pattern which may indeed contribute to a reduction in BC risk. This study is the first investigation of dietary effects on BC risk in Cyprus, a country whose population has traditionally adhered to the Mediterranean diet. Additional cohort studies of dietary patterns in our as well as in other Mediterranean populations are needed, to further clarify the protective effects of this diet on BC risk.

Acknowledgements

This work was supported by "Cyprus Research Promotion Foundation" grants 0104/13, 0104/17 and 32/2001, by EU-Europeaid grant CRIS 2009/223-507 and the Cyprus Institute of Neurology and Genetics. Further, we would like to thank all the study participants and acknowledge the help of the following cancer patient's organizations: the Pancyprian Association of Cancer Patients and Friends, Europa Donna Cyprus, the Cyprus Anticancer Society as well as Dr Vaios Partasides, director of the National breast cancer screening programme and his team. We also like to express our appreciation to Doctors Yiola Marcou, Panayiotis Papadopoulos, Maria Daniel and Simon Malas as well as to all the nurses and volunteers who provided valuable help towards the recruitment of the study participants. Lastly, we would like to thank dieticians Giorgos Loucaides, Rodoula Papalambrianou and Despo Michaelidou for the design of the FFQ as well as Stalo Lazarou for her advice and guidance concerning use of diet scores.

Author details

[1]Department of EM/Molecular Pathology, The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus. [2]Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK. [3]The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus. [4]Nutritional Epidemiology Unit, Istituto Nazionale Tumori, Milan, Italy. [5]Bank of Cyprus Oncology Center, Nicosia, Cyprus. [6]Department of Nursing, School of Health Sciences, Cyprus University of Technology, Limassol, Cyprus.

Authors' contributions

CAD performed the statistical analysis and interpretation of the dietary data and drafted the paper. AH MAL GL and KK have made substantial contributions to conception and design of the study and in the acquisition of data. They have also been involved in drafting the manuscript and revising it critically for important intellectual content. EK made substantial contributions to the acquisition of data. IN SS NM and PV were involved in drafting the manuscript and revising it critically for important intellectual content. All authors have given their approval for the submitted version.

Competing interests

The authors declare that they have no competing interests.

Received: 6 December 2011 Accepted: 23 March 2012
Published: 23 March 2012

References

1. GLOBOCAN Cancer Fact Sheets: Breast Cancer. [http://globocan.iarc.fr/factsheets/cancers/breast.asp].

2. Fung TT, Hu FB, McCullough ML, Newby PK, Willett WC, Holmes MD: **Diet quality is associated with the risk of estrogen receptor-negative breast cancer in postmenopausal women.** *J Nutr* 2006, **136**:466-472.
3. Trichopoulou A, Bamia C, Trichopoulos D: **Anatomy of health effects of Mediterranean diet: Greek EPIC prospective cohort study.** *BMJ* 2009, **338**: b2337.
4. Taylor EF, Burley VJ, Greenwood DC, Cade JE: **Meat consumption and risk of breast cancer in the UK Women's Cohort Study.** *Br J Cancer* 2007, **96**:1139-1146.
5. van der Hel OL, Peeters PHM, Hein DW, Doll MA, Grobbee DE, Ocké M, Bueno-de-Mesquita HB: **GSTM1 null genotype, red meat consumption and breast cancer risk (The Netherlands).** *Cancer Causes Control* 2004, **15**:295-303.
6. Egeberg R, Olsen A, Autrup H, Christensen J, Stripp C, Tetens I, Overvad K, Tjønneland A: **Meat consumption, N-acetyl transferase 1 and 2 polymorphism and risk of breast cancer in Danish postmenopausal women.** *Eur J Cancer Prev* 2008, **17**:39-47.
7. Suzuki R, Rylander-Rudqvist T, Ye W, Saji S, Adlercreutz H, Wolk A: **Dietary fiber intake and risk of postmenopausal breast cancer defined by estrogen and progesterone receptor status-a prospective cohort study among Swedish women.** *Int J Cancer* 2008, **122**:403-412.
8. Schutze M, Boeing H, Pischon T, Rehm J, Kehoe T, Gmel G, Olsen A, Tjonneland AM, Dahm CC, Overvad K, Clavel-Chapelon F, Boutron-Ruault M-C, Trichopoulou A, Benetou V, Zylis D, Kaaks R, Rohrmann S, Palli D, Berrino F, Tumino R, Vineis P, Rodriguez L, Agudo A, Sanchez M-J, Dorronsoro M, Chirlaque M-D, Barricarte A, Peeters PH, van Gils CH, Khaw K-T, Wareham N, Allen NE, Key TJ, Boffetta P, Slimani N, Jenab M, Romaguera D, Wark PA, Riboli E, Bergmann MM: **Alcohol attributable burden of incidence of cancer in eight European countries based on results from prospective cohort study.** *BMJ* 2011, **342**:d1584.
9. Sieri S, Krogh V, Muti P, Micheli A, Pala V, Crosignani P, Berrino F: **Fat and protein intake and subsequent breast cancer risk in postmenopausal women.** *Nutr Cancer* 2002, **42**:10-17.
10. Buckland G, Travier N, Gonzales CA, Lujan L, Agudo A, Trichopoulou A, Lagiou P, Trichopoulos D: **Adherence to the Mediterranean diet and risk of breast cancer in the EPIC cohort study.**.
11. Kant AK: **Indexes of overall diet quality: a review.** *J Am Diet Assoc* 1996, **96**:785-791.
12. Velie EM, Schairer C, Flood A, He J-P, Khattree R, Schatzkin A: **Empirically derived dietary patterns and risk of postmenopausal breast cancer in a large prospective cohort study.** *Am J Clin Nutr* 2005, **82**:1308-1319.
13. Couto E, Boffetta P, Lagiou P, Ferrari P, Buckland G, Overvad K, Dahm CC, Tjonneland A, Olsen A, Clavel-Chapelon F, Boutron-Ruault M-C, Cottet V, Trichopoulos D, Naska A, Benetou V, Kaaks R, Rohrmann S, Boeing H, von Ruesten A, Panico S, Pala V, Vineis P, Palli D, Tumino R, May A, Peeters PH, Bueno-de-Mesquita HB, Buchner FL, Lund E, Skeie G, Engeset D, Gonzalez CA, Navarro C, Rodriguez L, Sanchez M-J, Amiano P, Barricarte A, Hallmans G, Johansson I, Manjer J, Wirfart E, Allen NE, Crowe F, Khaw K-T, Wareham N, Moskal A, Slimani N, Jenab M, Romaguera D, Mouw T, Norat T, Riboli E, Trichopoulou A: **Mediterranean dietary pattern and cancer risk in the EPIC cohort.** *Br J Cancer* 2011, **104**:1493-1499.
14. Visioli F, Grande S, Bogani P, Galli C: **The role of antioxidants in the mediterranean diets: focus on cancer.** *Eur J Cancer Prev* 2004, **13**:337-343.
15. Ziegler RG: **Phytoestrogens and breast cancer.** *Am J Clin Nutr* 2004, **79**:183-184.
16. Hadjisavvas A, Loizidou MA, Middleton N, Michael T, Papachristoforou R, Kakouri E, Daniel M, Papadopoulos P, Malas S, Marcou Y, Kyriacou K: **An investigation of breast cancer risk factors in Cyprus: a case control study.** *BMC Cancer* 2010, **10**:447-447.
17. Psaltopoulou T, Kosti RI, Haidopoulos D, Dimopoulos M, Panagiotakos DB: **Olive oil intake is inversely related to cancer prevalence: a systematic review and a meta-analysis of 13800 patients and 23340 controls in 19 observational studies.** *Lipids Health Dis* 2011, **10**:127.
18. Michels KB, Mohllajee AP, Roset-Bahmanyar E, Beehler GP, Moysich KB: **Diet and breast cancer: a review of the prospective observational studies.** *Cancer* 2007, **109**:2712-2749.
19. Panagiotakos DB, Pitsavos C, Arvaniti F, Stefanadis C: **Adherence to the Mediterranean food pattern predicts the prevalence of hypertension, hypercholesterolemia, diabetes and obesity, among healthy adults; the accuracy of the MedDietScore.** *Prev Med* 2007, **44**:335-340.
20. Garrow J, James W, Ralph A: *Human Nutrition and Dietetics 10th Edition* London, UK: Churchhill Livingstone London.
21. Martínez-González MA, Fernández-Jarne E, Serrano-Martínez M, Wright M, Gomez-Gracia E: **Development of a short dietary intake questionnaire for the quantitative estimation of adherence to a cardioprotective Mediterranean diet.** *Eur J Clin Nutr* 2004, **58**:1550-1552.
22. Trichopoulou A, Bamia C, Trichopoulos D: **Mediterranean diet and survival among patients with coronary heart disease in Greece.** *Arch Intern Med* 2005, **165**:929-935.
23. Trichopoulou A, Bamia C, Lagiou P, Trichopoulos D: **Conformity to traditional Mediterranean diet and breast cancer risk in the Greek EPIC (European Prospective Investigation into Cancer and Nutrition) cohort.** *Am J Clin Nutr* 2010, **92**:620-625.
24. Cade JE, Taylor EF, Burley VJ, Greenwood DC: **Does the Mediterranean dietary pattern or the Healthy Diet Index influence the risk of breast cancer in a large British cohort of women [quest]?** *Eur J Clin Nutr* 2011, **65**:920-8.
25. Sofi F, Cesari F, Abbate R, Gensini GF, Casini A: **Adherence to Mediterranean diet and health status: meta-analysis.** *BMJ* 2008, **337**: a1344.
26. Ronco AL, de Stefani E, Aune D, Boffetta P, Deneo-Pellegrini H, Acosta G, Mendilaharsu M: **Nutrient patterns and risk of breast cancer in Uruguay.** *Asian Pac J Cancer Prev* 2010, **11**:519-524.
27. Murtaugh MA, Sweeney C, Giuliano AR, Herrick JS, Hines L, Byers T, Baumgartner KB, Slattery ML: **Diet patterns and breast cancer risk in Hispanic and non-Hispanic white women: the Four-Corners Breast Cancer Study.** *Am J Clin Nutr* 2008, **87**:978-984.
28. Cottet V, Touvier M, Fournier A, Touillaud MS, Lafay L, Clavel-Chapelon F, Boutron-Ruault M-C: **Postmenopausal breast cancer risk and dietary patterns in the E3N-EPIC prospective cohort study.** *Am J Epidemiol* 2009, **170**:1257-1267.
29. Sieri S, Krogh V, Pala V, Muti P, Micheli A, Evangelista A, Tagliabue G, Berrino F: **Dietary patterns and risk of breast cancer in the ORDET cohort.** *Cancer Epidemiol Biomarkers Prev* 2004, **13**:567-572.
30. Olsen A, Tjønneland A, Thomsen BL, Loft S, Stripp C, Overvad K, Møller S, Olsen JH: **Fruits and vegetables intake differentially affects estrogen receptor negative and positive breast cancer incidence rates.** *J Nutr* 2003, **133**:2342-2347.
31. Ravn-Haren G, Olsen A, Tjønneland A, Dragsted LO, Nexø BA, Wallin H, Overvad K, Raaschou-Nielsen O, Vogel U: **Associations between GPX1 Pro198Leu polymorphism, erythrocyte GPX activity, alcohol consumption and breast cancer risk in a prospective cohort study.** *Carcinogenesis* 2006, **27**:820-825.
32. Mattisson I, Wirfalt E, Johansson U, Gullberg B, Olsson H, Berglund G: **Intakes of plant foods, fibre and fat and risk of breast cancer - a prospective study in the Malmo Diet and Cancer cohort.** *Br J Cancer* 2004, **90**:122-127.
33. van Gils CH, Peeters PHM, Bueno-de-Mesquita HB, Boshuizen HC, Lahmann PH, Clavel-Chapelon F, Thiébaut A, Kesse E, Sieri S, Palli D, Tumino R, Panico S, Vineis P, Gonzalez CA, Ardanaz E, Sánchez M-J, Amiano P, Navarro C, Quirós JR, Key TJ, Allen N, Khaw K-T, Bingham SA, Psaltopoulou T, Koliva M, Trichopoulou A, Nagel G, Linseisen J, Boeing H, Berglund G, Wirfält E, Hallmans G, Lenner P, Overvad K, Tjønneland A, Olsen A, Lund E, Engeset D, Alsaker E, Norat T, Kaaks R, Slimani N, Riboli E: **Consumption of vegetables and fruits and risk of breast cancer.** *JAMA* 2005, **293**:183-193.
34. Smith-Warner SA, Spiegelman D, Yaun S-S, Adami H-O, Beeson WL, van den Brandt PA, Folsom AR, Fraser GE, Freudenheim JL, Goldbohm RA, Graham S, Miller AB, Potter JD, Rohan TE, Speizer FE, Toniolo P, Willett WC, Wolk A, Zeleniuch-Jacquotte A, Hunter DJ: **Intake of Fruits and Vegetables and Risk of Breast Cancer.** *JAMA: The J Am Med Assoc* 2001, **285**:769-776.
35. Zhang S, Hunter DJ, Hankinson SE, Giovannucci EL, Rosner BA, Colditz GA, Speizer FE, Willett WC: **A prospective study of folate intake and the risk of breast cancer.** *JAMA* 1999, **281**:1632-1637.
36. Sonestedt E, Gullberg B, Wirfält E: **Both food habit change in the past and obesity status may influence the association between dietary factors and postmenopausal breast cancer.** *Public Health Nutr* 2007, **10**:769-779.
37. Engeset D, Alsaker E, Lund E, Welch A, Khaw K-T, Clavel-Chapelon F, Thiébaut A, Chajès V, Key TJ, Allen NE, Amiano P, Dorronsoro M, Tjønneland A, Stripp C, Peeters PHM, van Gils CH, Chirlaque M-D, Nagel G, Linseisen J, Ocké MC, Bueno-de-Mesquita HB, Sacerdote C, Tumino R,

Ardanaz E, Sánchez M-J, Panico S, Palli D, Trichopoulou A, Kalapothaki V, Benetou V, Quirós JR, Agudo A, Overvad K, Bjerregaard L, Wirfält E, Schulz M, Boeing H, Slimani N, Riboli E: **Fish consumption and breast cancer risk. The European Prospective Investigation into Cancer and Nutrition (EPIC).** *Int J Cancer* 2006, **119**:175-182.

38. Li W, Ray RM, Lampe JW, Lin M-G, Gao DL, Wu C, Nelson ZC, Fitzgibbons ED, Horner N, Hu YW, Shannon J, Satia JA, Patterson RE, Stalsberg H, Thomas DB: **Dietary and other risk factors in women having fibrocystic breast conditions with and without concurrent breast cancer: a nested case-control study in Shanghai.** *China Int J Cancer* 2005, **115**:981-993.
39. Frazier AL, Ryan CT, Rockett H, Willett WC, Colditz GA: **Adolescent diet and risk of breast cancer.** *Breast Cancer Res* 2003, **5**:R59-R64.
40. Holmes MD, Colditz GA, Hunter DJ, Hankinson SE, Rosner B, Speizer FE, Willett WC: **Meat, fish and egg intake and risk of breast cancer.** *Int J Cancer* 2003, **104**:221-227.
41. Toniolo P, Riboli E, Shore RE, Pasternack BS: **Consumption of meat, animal products, protein, and fat and risk of breast cancer: a prospective cohort study in New York.** *Epidemiology* 1994, **5**:391-397.
42. Mills PK, Lawrence-Beeson W, Phillips RL, Fraser GE: **Dietary habits and breast cancer incidence among seventh-day adventists.** *Cancer* 1989, **64**:582-590.
43. Vatten LJ, Solvoll K, Løken EB: **Frequency of meat and fish intake and risk of breast cancer in a prospective study of 14,500 Norwegian women.** *Int J Cancer* 1990, **46**:12-15.
44. Gago-Dominguez M, Yuan J-M, Sun C-L, Lee H-P, Yu MC: **Opposing effects of dietary n-3 and n-6 fatty acids on mammary carcinogenesis: the Singapore Chinese Health Study.** *Br J Cancer* 2003, **89**:1686-1692.
45. Bank of Cyprus: *Bank of Cyprus Oncology Centre Registry* 2010.
46. Wu AH, Yu MC, Tseng C-C, Stanczyk FZ, Pike MC: **Dietary patterns and breast cancer risk in Asian American women.** *Am J Clin Nutr* 2009, **89**:1145-1154.

Pre-publication history

The pre-publication history for this paper can be accessed here:
http://www.biomedcentral.com/1471-2407/12/113/prepub

doi:10.1186/1471-2407-12-113
Cite this article as: Demetriou *et al.*: The mediterranean dietary pattern and breast cancer risk in Greek-Cypriot women: a case-control study. *BMC Cancer* 2012 **12**:113.

Submit your next manuscript to BioMed Central and take full advantage of:

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at
www.biomedcentral.com/submit

BioMed Central

APPENDIX B

The Mediterranean Diet and Breast Cancer Risk

The Mediterranean diet and breast cancer risk

Authors:

Christiana A. Demetriou[1,2], Andreas Hadjisavvas[1], Maria A. Loizidou[1], Paolo Vineis[2], Kyriacos Kyriacou[1]

[1]Department of Electron Microscopy and Molecular Pathology, The Cyprus Institute of Neurology and Genetics

[2]Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London

Author Contact Details:

Christiana A. Demetriou,	christianad@cing.ac.cy	(+357) 22392774
Andreas Hadjisavvas,	ahsavvas@cing.ac.cy	(+357) 22392739
Maria Loizidou	loizidou@cing.ac.cy	(+357) 22392633
Paolo Vineis	p.vineis@imperial.ac.cy	(+44) (0) 207 594 3372
Kyriacos Kyriacou	kyriacos@cing.ac.cy	(+357) 22392631

Institution Contact Details:

Department of Electron Microscopy and Molecular Pathology
The Cyprus Institute of Neurology and Genetics
P.O.Box 23462
1683 Nicosia
Cyprus
Telephone: +357 22 358 600
Telefax: +357 22 358 238

Abbreviated Title: Mediterranean diet and breast cancer risk

Abstract (for ScienceDirect)

There is ample evidence that the Mediterranean diet reduces breast cancer risk, despite the fact that findings across studies and populations have not been consistent. The protective effect of the Mediterranean diet is in line with the fact that some of its key components, such as fruit, vegetables, folate, and olive oil, have also been shown to have beneficial effects with regards to breast carcinogenesis.

The mechanisms through which these components are hypothesized to exert their protective effects include epigenetics and more specifically DNA methylation, anti-oxidant functions, biological response modification, anti-estrogenic, anti-proliferative, and anti-angiogenic activities.

Further research to investigate the effects of intensity and timing of exposure, and to elucidate the mechanistic pathways involved in the relationship is essential. The answers to these questions will provide biomarkers for prevention and early detection as well as targets for therapy, establishing the causality of the association and supporting adherence to the Mediterranean diet as a prevention strategy for breast cancer.

Key Words:

Mediterranean Diet, Mediterranean Diet Score, Breast Cancer, Vegetables, Olive Oil, Folate, DNA methylation, Exposome, Biomarkers.

INTRODUCTION

Breast cancer incidence trends show widespread variation as evident in Figure 1. The sharp increase in incidence during the past decades suggests that environmental exposures interact with genetics, making breast cancer a truly multifactorial disease. Several reproductive factors as well as environmental exposures have been associated with breast cancer risk [1].

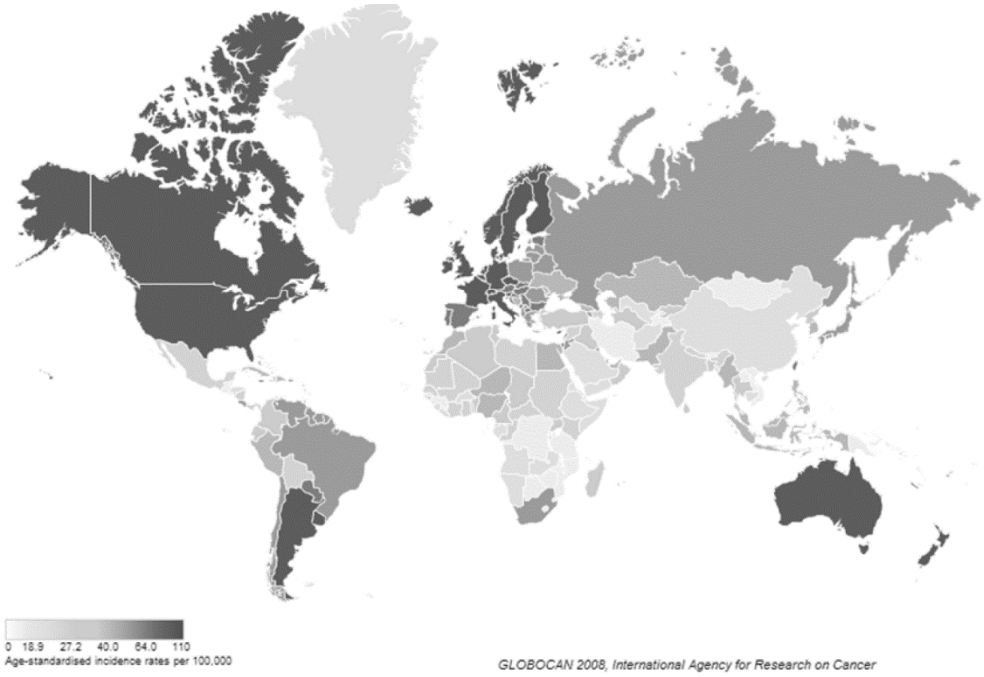


Figure 1 – Worldwide Map of Breast Cancer Age-adjusted Incidence Rates in 2008

World map shows widespread variation in incidence of breast cancer. Darker colours indicate higher incidence of breast cancer. Figure used with permission from The International Agency for Research on Cancer; Lyon, France © 2010.

Diet has long been suspected to affect breast cancer risk and specific food groups and dietary nutrients have been investigated in a number of prospective studies to elucidate whether there is an association between diet components and breast cancer risk.

However, when investigating the effects of diet on a particular health outcome, one needs to consider that individuals do not consume single foods, but combinations of several foods that contain both nutrient and non-nutrient substances. For this reason, it is useful to

examine patterns of nutrient intake that express several related aspects of dietary intake concurrently [2].

Several studies investigated different dietary patterns for their effects on breast cancer risk, and the Mediterranean diet has been one of the most researched patterns. The Mediterranean diet encompasses all the beneficial dietary characteristics presented in nutritional studies and thus it may offer protection from breast cancer [3]. The combination and range of foods included in this diet pattern, provide plenty of anti-oxidants such as flavonoids, carotenoids, and vitamins, lots of phytochemicals including phytoestrogens, sufficient quantities of fiber, adequate folate, and a favourable fatty acid profile [4,5].

These nutrients have been linked to mechanisms of carcinogenesis and have been found to confer protective effects [6] on total cancer incidence and more relevantly on breast cancer incidence. This might explain the repeatedly observed lower breast cancer incidence in Southern European countries, compared to the rest of Europe as well as North America (Figure 2).

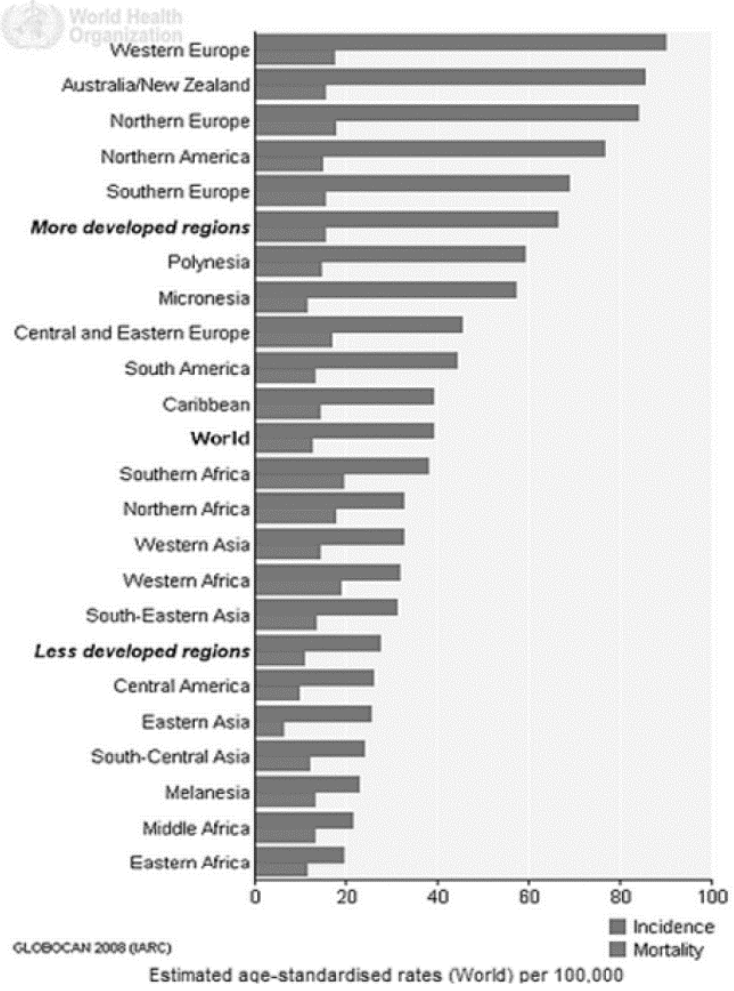


Figure 2 – Breast Cancer Incidence and Mortality Worldwide in 2008

The figure displays age-standardized breast cancer rates for incidence and mortality. It is evident that rates in Southern Europe, where most Mediterranean countries lay, are lower than rates in Northern, Central and Western Europe. Figure used with permission from The International Agency for Research on Cancer; Lyon, France © 2010.

THE MEDITERRANEAN DIETARY PATTERN AND BREAST CANCER RISK: EPIDEMIOLOGICAL EVIDENCE

Interestingly, few studies have been carried out on Mediterranean populations in order to investigate the populations' adherence to the Mediterranean diet in the age of the widespread Westernized lifestyle. In addition, currently there are few studies from other countries that investigate the possible health benefits of the diet on breast cancer risk. The prospective studies investigating the association between the Mediterranean diet pattern and breast cancer risk are summarized in Table 1.

Recently a study in the EPIC cohort demonstrated that higher adherence to a Mediterranean dietary pattern – as assessed using the Mediterranean diet score developed by Trichopoulou [7] – is associated with a decreased cancer risk both in Mediterranean and in non-Mediterranean countries (HR: 0.96, 95% CI 0.96-0.98 for a two point increment in the score). This analysis did not identify any of the components of the score as having a predominant effect supporting the idea that it is the combination and range of nutrient and non-nutrient components of the diet, that confer the protective effect [4]. In a more focused analysis of the EPIC cohort with respect to breast cancer risk, adherence to the Mediterranean diet pattern – assessed using an adaptation of the Trichopoulou score – was found to confer a small decrease in total and post-menopausal breast cancer risk. The protective effect became stronger when no consumption of alcohol – instead of moderate –was treated as most beneficial and was more pronounced for ER-/PR- postmenopausal tumours. No association for premenopausal women was found [5].

A study on the Greek EPIC cohort has also demonstrated a marginally significant inverse association between adherence to a Mediterranean diet pattern (defined a priori) and postmenopausal breast cancer [8]. Similarly, in a case-control study of Asian American women, high adherence to the Mediterranean diet pattern was found to be associated with reduced breast cancer risk [3].

However, Cade et al. [9] that examined the adherence to a Mediterranean diet pattern in a British cohort did not yield any significant results, except a non-significant trend towards decreased risk with increasing adherence in pre-menopausal women. This cohort was designed to include women with different dietary patterns and at extremes of intake, so as

to enable comparisons and maximize the potential of successful identification of any interactions (greater contrasts reduce the effect of measurement error). Similarly, Couto et al. did not find adherence to a Mediterranean diet pattern to be associated with reduced breast cancer risk, in the Swedish Women's Lifestyle and Health cohort [10] . Fung et al. [11] investigated postmenopausal breast cancer risk in the Nurses' Health Study, but found no significant association between adherence to the Mediterranean diet and breast cancer. However, results for ER- breast cancer approached significance. In addition, Nkondjock et al. [12], in a case control study did not find adherence to the Mediterranean diet to influence BRCA-associated breast cancer risk.

A meta analysis of 6 prospective studies with a total of 512,366 subjects followed for a period ranging from eight to 18 years, showed that adherence to the Mediterranean diet is inversely associated with total cancer incidence and mortality, but there were no results specific to breast cancer [13].

It is argued that results from Northern European countries may differ from results from Mediterranean populations, because essential components of the Mediterranean diet may be consumed in different forms and proportions in Northern European countries [9]. Therefore, instead of utilizing a priori scores to define a population's diet according to its adherence to the Mediterranean diet pattern, a number of studies utilised factor analysis to derive diet patterns specific to the population under study.

Table 1 – Cohort Studies investigating the association between adherence to the Mediterranean Diet and Breast Cancer Risk

Studies listed in chronological order. Two studies show a significant reduction in postmenopausal breast cancer risk with increasing adherence to the Mediterranean Diet.

Author, Year	Study, Country	Participants	Exposure Assessment, Exposure Category Comparison	Effect Estimate (95% Confidence Interval)
Fung, 2006[11]	Nurses' Health Study, USA	3,580 cases of postmenopausal breast cancer (2,367 ER+ and 575 ER- cancers)	alternate Mediterranean Diet Score (aMed), Quartile 5 vs. Quartile 1	ER+ cancers: 1.05 (0.91, 1.18) ER- cancers: 0.79 (0.60, 1.03)
Trichopoulou, 2010[8]	EPIC Cohort, Greece	14,807 women (240 incident breast cancer cases)	Mediterranean diet score (Range 0-9), Two-point score increment	All women: 0.88 (0.75, 1.03) Premenopausal women: 1.01 (0.80, 1.28) Postmenopausal women 0.78 (0.62, 0.98)
Cade, 2011[9]	UK Women's Cohort Study, United Kingdom	33,731 women (828 incident breast cancer cases)	10 component Mediterranean diet score (Range: 0-10), High (7-10) vs. Low (0-2) score	All women: 0.96 (0.60, 1.32) Premenopausal women: 0.65 (0.42, 1.02) Postmenopausal women: 1.30 (0.83, 2.05)
Buckland, 2012[5]	EPIC Cohort, 10 European Countries	142,605 men 335,873 women (9,009 postmenopausal and 1,216 premenopausal incident breast cancer cases)	adapted relative Mediterranean diet (arMED) score excluding alcohol (Range 0-16), High (10-16) vs. Low (0-5) score	All women: 0.94 (0.88, 1.00) Premenopausal women: 0.97 (0.81, 1.15) Postmenopausal women: 0.93 (0.87, 0.99)
Couto, 2013[10]	Swedish Women's Lifestyle and Health cohort study, Sweden	49,258 women (1,278 incident breast cancers)	Mediterranean diet score (Range 0-9), Two-point score increment	All women: 1.08 (1.00–1.15) Premenopausal women: 1.10 (1.01–1.21) Postmenopausal women: 1.02 (0.91–1.15)

Velie et al. [2] found that a vegetable-fish/poultry-fruit diet pattern was not significantly correlated with breast cancer. However, other studies yielded significant results. Ronco et al. [14] investigated the relationship between food patterns and breast cancer risk in Uruguay and

found that adherence to a food pattern rich in anti-oxidants – associated with white meat, fruit, vegetables – was protective. Murtaugh et al. [15] in their study of Hispanic and non-Hispanic white women evidenced that a dietary pattern labeled as Mediterranean, due to its similarities with the Mediterranean diet was statistically significant and inversely associated with breast cancer. Furthermore, in the French subgroup of the EPIC cohort, a "healthy/Mediterranean" diet pattern (essentially vegetables, fruits, seafood, olive oil, and sunflower oil) was significantly inversely associated with breast cancer risk, especially for ER+/PR- cancers [16]. Lastly, Demetriou et al. demonstrated a reduced breast cancer risk, among Mediterranean women adhering to a Principal Component Analysis (PCA) derived dietary pattern rich in vegetables, fruit, fish, and legumes – all important components of the Mediterranean diet [17]. On the contrary, Bessaud et al. in a case control study in Southern France showed that high intake of raw vegetables decreased breast cancer risk [18], but a PCA derived Mediterranean dietary pattern was not associated with breast cancer risk [19].

From the current evidence it becomes apparent that the definition of the Mediterranean diet used and whether this is defined a-priori or a-posteriori, or even the specific score used to assess adherence to the dietary pattern, as well as the population under study can affect the results of studies. It is not surprising then, that researchers are advocating for more studies to further elucidate the relationship between a Mediterranean diet pattern and breast cancer risk. Studies in Mediterranean countries are further warranted since it can be argued that countries in the Mediterranean region adhere to a more “pure” Mediterranean diet and they also exhibit lower prevalence of breast cancer, compared to countries following a more westernized type of diet (Figure 2). Such studies may explain, at least in part, the repeatedly observed lower breast cancer incidence in Southern European countries, most of which border the Mediterranean Sea.

COMPONENTS OF THE MEDITERRANEAN DIET AND BREAST CANCER RISK: EPIDEMIOLOGICAL EVIDENCE

Given the variation in results depending on the score used to define Mediterranean diet adherence and the population under study, a more dependable approach would be to investigate the components of the Mediterranean diet for their impact on breast cancer risk. Indeed, there is ample experimental evidence that fruit and vegetables, as well as olive oil, which are essential components of the Mediterranean diet, are independently associated with reduced breast cancer incidence in a few high powered, prospective studies, particularly with certain subgroups of breast cancer.

Shannon et al. (2005)[20], in a nested case control study in China, evidenced that women in the highest quartile of fruit and vegetable consumption had a significantly decreased risk of breast cancer. However, no single botanical group could explain this association suggesting that it is in the combination and variety of intake that its benefits lie. A decreasing risk of breast cancer with increasing intake of vegetables rich in vitamins A and C was also observed, in another nested case-control study but the trend was not significant [21]. Zhang et al. (1999)[22] in a report from the Nurses' Health Study, showed that intake of 5 or more portions of fruit and vegetables daily was associated with decreased breast cancer risk compared to an intake of less than 2 portions daily, in premenopausal women. This association was strongest in women with a family history of breast cancer or in women who consumed more than 15g of alcohol per day. No association was observed in postmenopausal women. Olsen et al. (2003)[23] in the Danish Diet, Cancer and Health Study did not find an association between consumption of fruit and vegetables with overall cancer incidence. They did, however, observe that for ER+ tumours only, intake of fruit and vegetables increased breast cancer risk, with borderline significance. A more recent case control study nested in this cohort, showed no association between intake of fruit and vegetables and postmenopausal breast cancer[24]. In the Malmo Diet and Cancer Cohort, no association between intake of any plant groups and postmenopausal breast cancer was evidenced[25]. In an updated analysis of the Malmo cohort data, an overall association between intake of fruit, berries, and vegetables with breast cancer was again not demonstrated. However, in postmenopausal women with a BMI less than 27kg/m^2 and in postmenopausal women not changing their diet before recruitments, there was a significant reduction in breast cancer risk with increased intake of fruit, berries and vegetables[26]. In the

same regard, analysis of the Greek cohort of the EPIC study evidenced that a high consumption of vegetables and fruits was one of the most beneficial components, for the protective effect of the Mediterranean diet on survival [27].

Folate is a key nutrient in fruit and vegetables and folate has also been shown to be associated with reduced breast cancer risk, especially in postmenopausal women with increased alcohol intake. In the ACS Cancer Prevention Study II [28], the Iowa Health Study [29] and in the Malmo Diet and Cancer Cohort [25] folate intake was not significantly associated with postmenopausal breast cancer risk. On the contrary, Rohan et al. [30] analysing data from the Canadian National Breast Screening Study and Zhang et al. [22] analysing data from the Nurses' Health Study, both demonstrated statistically significant protective effect of folate intake on postmenopausal breast cancer risk, in women whose alcohol consumption exceeded 15g of alcohol per day. A meta-analysis of four of these studies [25,28–30] performed by The Breast Cancer Report 2006, part of the WCRF/AICR Systematic Literature Review Continuous Update Report yielded a non-statistically significant RR of 0.90.

However, more recent reports are more consistent in showing a protective effect of folate on postmenopausal breast cancer risk. The Malmo Diet and Cancer Cohort [31] demonstrated a significant reduction in postmenopausal breast cancer risk with increasing dietary folate intake and a statistically significant decrease in risk was also demonstrated among postmenopausal women in the French subgroup of the EPIC cohort [32]. A non-significant protective effect of folate intake on postmenopausal breast cancer was also demonstrated by two reports from two cohort studies - Stolzenberg-Solomon et al. [33] from the PLCO Cancer Screening Trial Cohort and Tjonneland et al. [34] from the Diet, Cancer and Health Cohort.

In a published meta-analysis, a 200mg/day increase in folate intake (dietary only or total), was not associated with postmenopausal breast cancer risk in prospective studies. However, in case-control studies there was a statistically significant decrease in risk for dietary folate intake but not for total intake [35].

The evidence above is not sufficient to support a protective effect of folate intake on breast cancer risk per se but the findings of a protective effect in those individuals with alcohol intakes above certain thresholds warrant further investigation.

Regarding olive oil and breast cancer risk, in the European countries participating in EPIC, an association was not evidenced [36], but in a systematic review and meta-analysis of 13,800 patients and 23,340 controls in 19 observational studies, olive oil intake was associated with lower odds of developing breast cancer [37]. Similarly, in another report including both human and animal studies, diets rich in extra virgin olive oil (EVOO) were found to exert a negative modulatory effect on experimental breast cancer. In addition, EVOO was shown to induce different molecular changes in tumours, such as in the composition of cell membranes, activity of signalling proteins and gene expression [38].

THE PROTECTIVE EFFECT OF THE MEDITERRANEAN DIET: MECHANISTIC EVIDENCE

The inverse association between the Mediterranean diet and breast cancer risk can be biologically explained by its high content in nutrients shown to confer a protective effect with regards to carcinogenesis pathways.

Folate, found largely in fruit and vegetables, is a key player in the one carbon metabolism pathway. This pathway is not only responsible for the methylation of biological molecules such as DNA and RNA, it is also essential for DNA synthesis and repair. This is because folate derived one carbon groups, are needed for the de novo synthesis of pyrimidine, thymidylate, as well as the purines [39]. There is evidence that folate deficiency may lead to global hypomethylation [40] which is an event frequently observed in breast tumours, and is associated with activation of oncogenes as well as chromosomal instability through repetitive element activation [41]. In addition, given folate's role in DNA integrity, synthesis, and repair, folate deficiency may compromise the reliability of these processes leading to tolerance in DNA damage which is also associated with carcinogenesis[42].

The Mediterranean diet is also rich in carotenoids, such as retinol, retinyl pulminate, α-carotene, β-carotene, β-cryptoxanthin, lutein and lycopene. These substances have been shown to decrease proliferation and promote cell differentiation of epithelial cells in culture [43,44]. In addition, carotenoids have anti-oxidant function and can protect macromolecules such as DNA from damage from Reactive Oxygen Species (ROS)[42].

In addition to folate and carotenoids, flavonoids are also abundant in the Mediterranean diet. Flavonoids, such as catechins, are polyphenolic compounds found in plants, and have been shown to be biological response modifiers. High flavonoid intake has been consistently shown to decrease breast cancer risk, and this protective effect might stem from their effect on ovarian sex hormone synthesis, or their other biological activities such as antioxidant, antiproliferative, and antiangiogenic activities (summarised in [45]). In addition, flavonoids have been shown to inhibit the effects of cytokines, growth factors, and several other enzymes involved in carcinogenic processes (summarised in [45]).

Phytoestrogens, plant-derived xenoestrogens, are found in large amounts in the Mediterranean diet because of its high content in fruit, vegetables, and fibre. Phytoestrogens have a similar molecular structure to oestrogens which allows them to have oestrogenic or anti-oestrogenic effects by binding to estrogen receptors [46]. Phytoestrogens

can also modulate the bioavailability of sex hormones by stimulating or inhibiting the action of enzymes such as the Sex Hormone Binding Globulin (SHBG)[47]. The anti-oestrogenic effects of phytoestrogens have been shown to confer a protective effect for breast cancer, however the evidence has not been always consistent [48].

Lastly, olive oil and especially virgin olive oil intake is common among Mediterranean populations. In fact, in the Mediterranean diet, olive oil is the principal source of dietary fat, so that there is a high polyunsaturated fatty acid (PUFA) to saturated fatty acid (SFA) intake ratio. This favourable fatty acid ratio helps maintain a healthy body weight which is protective with regards to breast carcinogenesis [38]. In addition, EVOO alters the morphology and molecular characteristics of tumours, such as cell membrane composition, signaling protein activity, and gene expression [38], in a way that suggests also possible protective effects.

These chemo-preventive mechanisms involving micronutrients abundant in the Mediterranean diet provide biological plausibility to the epidemiological evidence linking the Mediterranean diet to a reduced breast cancer risk.

Investigating the effect of the Mediterranean diet on breast cancer risk, carries with it many challenges. Firstly, despite the complexity of diet itself, it is only a small fraction of the totality of every exposure that one is exposed to from conception to death – defined as the exposome [49]. The exposome includes internal exposures such as metabolism, hormones, and inflammation, general external exposures such as education and financial status, as well as specific external exposures such as infectious agents, lifestyle factors, and diet [49] (Figure 3). It is also true that these three domains of the exposome are not mutually exclusive. For example, a high fat diet during childhood which is considered a specific external exposure can upset [59] the hormonal balance which is an internal exposure. In addition, education and socioeconomic status can influence one's dietary choices.

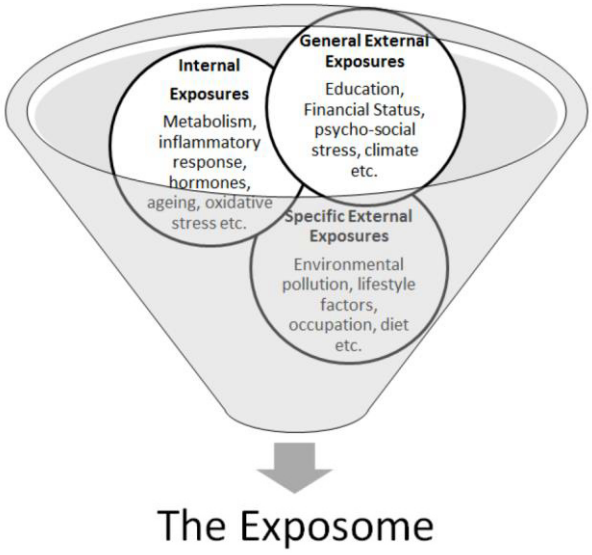


Figure 3 – Diagrammatical representation of the three domains of the exposome

The exposome consists of the totality of internal, general external and specific external exposures which interact in an additive and complementary nature as described in [49].

When studying exposures, one should not only consider the interplay between the three domains of the exposome and the countless components of each, but also the dynamic nature of the exposome. For example, diet in utero is far and wide different from diet during

childhood and adolescence, and diet in adolescence often differs dramatically from diet in adult life and middle age. To further complicate the picture, some exposures tend to have a greater effect during developmental windows when internal systems such as inflammatory response, epigenetics, and hormone balance are more plastic [42,49]. For example, as illustrated in Figure 4, breast tissue undergoes different phases of vulnerability to external exposures though different life stages.

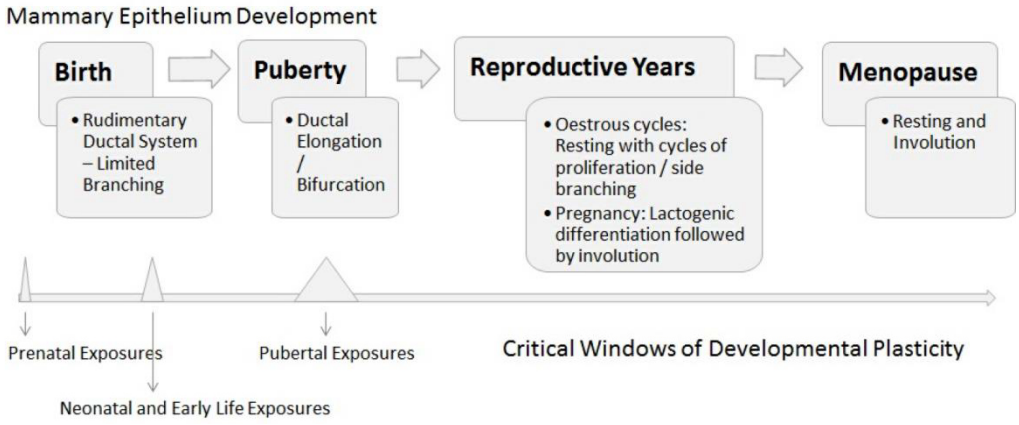


Figure 4 – Mammary Epithelium Development and Critical Windows of Developmental Plasticity

Mammary tissue development consists of different timeframes of rudimentary or extensive proliferation, differentiation, resting, and involution. Mammary tissue is particularly vulnerable to external exposures during periods of proliferation, hence prenatal, neonatal, and pubertal periods are considered critical windows of developmental plasticity. Exposures during these time frames are likely to have a greater effect on the proliferating tissue, than exposures during periods when the tissue is resting. Figure adapted from [42] with permission from Cambridge University Press © 2012 .

As a result, in order to get an accurate picture of the interplay between the Mediterranean diet and breast cancer, precise assessment of the exposure at different time-frameworks to determine when diet is more aetiologically relevant is warranted. Furthermore, different metrics of exposure such as duration of adherence in addition to degree of adherence need to be investigated. Given all these challenges it is not surprising that studies fail to produce consistent evidence on the association between the Mediterranean diet with its related components and breast cancer risk.

Of great importance in the endeavour to understand the relationship between the Mediterranean diet and breast cancer risk are the studies investigating the mechanistic evidence which links components of the diet and breast carcinogenesis. Elucidation of the mechanisms involved will aid in the identification of biomarkers of the exposure as well as biomarkers of the disease and these findings will in turn add weight to the association helping researchers assign causality. Vineis and Perera [50] propose the use of complementary study designs to establish new biomarkers that lie on the causal pathway between exposure and disease. For example, identification of biomarkers that in prospective studies are shown to be associated with adherence to the Mediterranean diet and evidence from nested case-control studies that these same biomarkers are predictive of breast cancer risk, will be sufficient to classify these biomarkers as part of the mechanistic pathway linking exposure to disease and to regard this association as causal (Figure 4). This approach, termed "Meet-in-the-Middle", can open up new avenues for establishing causal relationships which will then shed light on possible targets for prevention and treatment. Thus, "Meet-in-the-Middle" will be invaluable in future research efforts for establishing the beneficial effects of the Mediterranean diet on disease risk.

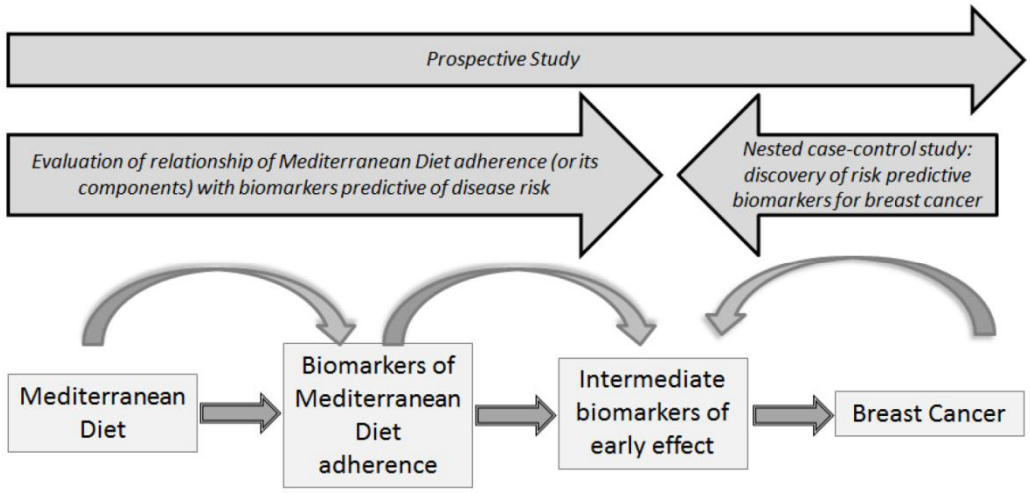


Figure 5 – Meet-in-the-Middle

The figure demonstrates the use of complementary study designs to identify biomarkers that lie on the causal pathway between exposure and disease.

CONCLUSIONS

The pool of evidence above, justifies the promotion of the Mediterranean diet as a healthy dietary choice for the reduction of a woman's breast cancer risk. It also calls for more research to further elucidate the degree of protection conferred from this dietary pattern and the mechanisms involved in risk reduction. This is important because even if the reduction in risk conferred by a strong adherence to the Mediterranean dietary pattern is small, given the high incidence of breast cancer, then even a small reduction in risk would translate into a large number of breast cancer cases prevented.

Summary Points:

- There is ample evidence that the Mediterranean diet reduces breast cancer risk but results across studies and across populations have not been consistent
- Components of the Mediterranean diet – such as vegetables, fruit, folate, and olive oil – have been shown to exert beneficial effects on breast cancer risk
- The mechanisms through which Mediterranean diet components exert their protective effects include DNA methylation, anti-oxidant functions, biological response modification, anti-oestrogenic, anti-proliferative, and anti-angiogenic activities
- In the investigations of the relationship between Mediterranean diet adherence and breast cancer risk, factors such as intensity and timing of exposure are key and need to be considered
- Further research, especially in Mediterranean populations, is essential to establish the causality of the association and support adherence to the Mediterranean diet as a preventive strategy for breast cancer

References:

1 McPherson K, Steel CM, Dixon JM. ABC of breast diseases. Breast cancer-epidemiology, risk factors, and genetics. *BMJ* 2000; **321**: 624–8.

2 Velie EM, Schairer C, Flood A, He J-P, Khattree R, Schatzkin A. Empirically derived dietary patterns and risk of postmenopausal breast cancer in a large prospective cohort study. *Am J Clin Nutr* 2005; **82**: 1308–19.

3 Wu AH, Yu MC, Tseng C-C, Stanczyk FZ, Pike MC. Dietary patterns and breast cancer risk in Asian American women. *Am J Clin Nutr* 2009; **89**: 1145–54.

4 Couto E, Boffetta P, Lagiou P, *et al.* Mediterranean dietary pattern and cancer risk in the EPIC cohort. *Br J Cancer* 2011; **104**: 1493–9.

5 Buckland G, Travier N, Cottet V, *et al.* Adherence to the mediterranean diet and risk of breast cancer in the EPIC cohort study. *International Journal of Cancer* 2012. doi:10.1002/ijc.27958.

6 Visioli F, Grande S, Bogani P, Galli C. The role of antioxidants in the mediterranean diets: focus on cancer. *Eur J Cancer Prev* 2004; **13**: 337–43.

7 Trichopoulou A, Bamia C, Trichopoulos D. Mediterranean diet and survival among patients with coronary heart disease in Greece. *Arch Intern Med* 2005; **165**: 929–35.

8 Trichopoulou A, Bamia C, Lagiou P, Trichopoulos D. Conformity to traditional Mediterranean diet and breast cancer risk in the Greek EPIC (European Prospective Investigation into Cancer and Nutrition) cohort. *Am J Clin Nutr* 2010; **92**: 620–5.

9 Cade JE, Taylor EF, Burley VJ, Greenwood DC. Does the Mediterranean dietary pattern or the Healthy Diet Index influence the risk of breast cancer in a large British cohort of women[quest]. *Eur J Clin Nutr* 2011.http://dx.doi.org/10.1038/ejcn.2011.69 (accessed 24 May2011).

10 Couto E, Sandin S, Löf M, Ursin G, Adami H-O, Weiderpass E. Mediterranean Dietary Pattern and Risk of Breast Cancer. *PLoS ONE* 2013; **8**: e55374.

11 Fung TT, Hu FB, McCullough ML, Newby PK, Willett WC, Holmes MD. Diet quality is associated with the risk of estrogen receptor-negative breast cancer in postmenopausal women. *J Nutr* 2006; **136**: 466–72.

12 Nkondjock A, Ghadirian P. Diet quality and BRCA-associated breast cancer risk. *Breast Cancer Res Treat* 2007; **103**: 361–9.

13 Sofi F, Cesari F, Abbate R, Gensini GF, Casini A. Adherence to Mediterranean diet and health status: meta-analysis. *BMJ* 2008; **337**: a1344–a1344.

14 Ronco AL, de Stefani E, Aune D, *et al.* Nutrient patterns and risk of breast cancer in Uruguay. *Asian Pac J Cancer Prev* 2010; **11**: 519–24.

15 Murtaugh MA, Sweeney C, Giuliano AR, *et al.* Diet patterns and breast cancer risk in Hispanic and non-Hispanic white women: the Four-Corners Breast Cancer Study. *The American Journal of Clinical Nutrition* 2008; **87**: 978 –984.

16 Cottet V, Touvier M, Fournier A, *et al.* Postmenopausal breast cancer risk and dietary patterns in the E3N-EPIC prospective cohort study. *Am J Epidemiol* 2009; **170**: 1257–67.

17 Demetriou CA, Hadjisavvas A, Loizidou MA, *et al.* The mediterranean dietary pattern and breast cancer risk in Greek-Cypriot women: a case-control study. *BMC Cancer* 2012; **12**: 113.

18 Bessaoud F, Daurès J-P, Gerber M. Dietary factors and breast cancer risk: a case control study among a population in Southern France. *Nutr Cancer* 2008; **60**: 177–87.

19 Bessaoud F, Tretarre B, Daurès J-P, Gerber M. Identification of dietary patterns using two statistical approaches and their association with breast cancer risk: a case-control study in Southern France. *Ann Epidemiol* 2012; **22**: 499–510.

20 Shannon J, Ray R, Wu C, *et al.* Food and Botanical Groupings and Risk of Breast Cancer: A Case-Control Study in Shanghai, China. *Cancer Epidemiology Biomarkers & Prevention* 2005; **14**: 81 –90.

21 Rohan TE, Howe GR, Friedenreich CM, Jain M, Miller AB. Dietary fiber, vitamins A, C, and E, and risk of breast cancer: a cohort study. *Cancer Causes Control* 1993; **4**: 29–37.

22 Zhang S, Hunter DJ, Forman MR, *et al.* Dietary Carotenoids and Vitamins A, C, and E and Risk of Breast Cancer. *Journal of the National Cancer Institute* 1999; **91**: 547 –556.

23 Olsen A, Tjønneland A, Thomsen BL, *et al.* Fruits and vegetables intake differentially affects estrogen receptor negative and positive breast cancer incidence rates. *J Nutr* 2003; **133**: 2342–7.

24 Ravn-Haren G, Olsen A, Tjønneland A, *et al.* Associations between GPX1 Pro198Leu polymorphism, erythrocyte GPX activity, alcohol consumption and breast cancer risk in a prospective cohort study. *Carcinogenesis* 2006; **27**: 820–5.

25 Mattisson I, Wirfält E, Johansson U, Gullberg B, Olsson H, Berglund G. Intakes of plant foods, fibre and fat and risk of breast cancer--a prospective study in the Malmö Diet and Cancer cohort. *Br J Cancer* 2004; **90**: 122–7.

26 Sonestedt E, Gullberg B, Wirfält E. Both food habit change in the past and obesity status may influence the association between dietary factors and postmenopausal breast cancer. *Public Health Nutr* 2007; **10**: 769–79.

27 Trichopoulou A, Bamia C, Trichopoulos D. Anatomy of health effects of Mediterranean diet: Greek EPIC prospective cohort study. *BMJ* 2009; **338**: b2337–b2337.

28 Feigelson HS, Jonas CR, Robertson AS, McCullough ML, Thun MJ, Calle EE. Alcohol, folate, methionine, and risk of incident breast cancer in the American Cancer Society Cancer Prevention Study II Nutrition Cohort. *Cancer Epidemiol Biomarkers Prev* 2003; **12**: 161–4.

29 Sellers TA, Grabrick DM, Vierkant RA, *et al.* Does folate intake decrease risk of postmenopausal breast cancer among women with a family history? *Cancer Causes Control* 2004; **15**: 113–20.

30 Rohan TE, Jain M, Howe GR, Miller AB. Alcohol consumption and risk of breast cancer: a cohort study. *Cancer Causes Control* 2000; **11**: 239–47.

31 Ericson U, Sonestedt E, Gullberg B, Olsson H, Wirfält E. High folate intake is associated with lower breast cancer incidence in postmenopausal women in the Malmö Diet and Cancer cohort. *Am J Clin Nutr* 2007; **86**: 434–43.

32 Lajous M, Lazcano-Ponce E, Hernandez-Avila M, Willett W, Romieu I. Folate, vitamin B(6), and vitamin B(12) intake and the risk of breast cancer among Mexican women. *Cancer Epidemiol Biomarkers Prev* 2006; **15**: 443–8.

33 Stolzenberg-Solomon RZ, Chang S-C, Leitzmann MF, *et al.* Folate intake, alcohol use, and postmenopausal breast cancer risk in the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial. *Am J Clin Nutr* 2006; **83**: 895–904.

34 Tjønneland A, Christensen J, Olsen A, *et al.* Folate intake, alcohol and risk of breast cancer among postmenopausal women in Denmark. *Eur J Clin Nutr* 2006; **60**: 280–6.

35 Larsson SC, Giovannucci E, Wolk A. Folate and risk of breast cancer: a meta-analysis. *J Natl Cancer Inst* 2007; **99**: 64–76.

36 Buckland G, Travier N, Agudo A, *et al.* Olive oil intake and breast cancer risk in the Mediterranean countries of the European Prospective Investigation into Cancer and Nutrition study. *Int J Cancer* 2012; **131**: 2465–9.

37 Psaltopoulou T, Kosti RI, Haidopoulos D, Dimopoulos M, Panagiotakos DB. Olive oil intake is inversely related to cancer prevalence: a systematic review and a meta-analysis of 13800 patients and 23340 controls in 19 observational studies. *Lipids Health Dis* 2011; **10**: 127.

38 Escrich E, Moral R, Solanas M. Olive oil, an essential component of the Mediterranean diet, and breast cancer. *Public Health Nutr* 2011; **14**: 2323–32.

39 Duthie SJ. Folate and cancer: how DNA damage, repair and methylation impact on colon carcinogenesis. *J Inherit Metab Dis* 2010. doi:10.1007/s10545-010-9128-0.

40 Chen J, Gammon MD, Chan W, *et al.* One-carbon metabolism, MTHFR polymorphisms, and risk of breast cancer. *Cancer Res* 2005; **65**: 1606–14.

41 Jovanovic J, Rønneberg JA, Tost J, Kristensen V. The epigenetics of breast cancer. *Mol Oncol* 2010; **4**: 242–54.

42 Teegarden D, Romieu I, Lelièvre SA. Redefining the impact of nutrition on breast cancer incidence: is epigenetics involved? *Nutr Res Rev* 2012; **25**: 68–95.

43 Rabi T, Bishayee A. Terpenoids and breast cancer chemoprevention. *Breast Cancer Res Treat* 2009; **115**: 223–39.

44 Okuno M, Kojima S, Matsushima-Nishiwaki R, *et al.* Retinoids in cancer chemoprevention. *Curr Cancer Drug Targets* 2004; **4**: 285–98.

45 Hui C, Qi X, Qianyong Z, Xiaoli P, Jundong Z, Mantian M. Flavonoids, flavonoid subclasses and breast cancer risk: a meta-analysis of epidemiologic studies. *PLoS ONE* 2013; **8**: e54318.

46 Humfrey CD. Phytoestrogens and human health effects: weighing up the current evidence. *Nat Toxins* 1998; **6**: 51–9.

47 Pino AM, Valladares LE, Palma MA, Mancilla AM, Yáñez M, Albala C. Dietary isoflavones affect sex hormone-binding globulin levels in postmenopausal women. *J Clin Endocrinol Metab* 2000; **85**: 2797–800.

48 Miller PE, Snyder DC. Phytochemicals and cancer risk: a review of the epidemiological evidence. *Nutr Clin Pract* 2012; **27**: 599–612.

49 Wild CP. The exposome: from concept to utility. *Int J Epidemiol* 2012. doi:10.1093/ije/dyr236.

50 Vineis P, Perera F. Molecular epidemiology and biomarkers in etiologic cancer research: the new in light of the old. *Cancer Epidemiol Biomarkers Prev* 2007; **16**: 1954–65.

APPENDIX C

List of Genes whose promoter aberrant methylation is implicated in breast cancer

(Identified from PUBMETH and extensive literature review of the years 2007 – May 2013)

GENE NAME	REFERENCES (PUBMETH + Literature Review)	REFERENCES (TOTAL)	AVERAGE METHYLATION FREQUENCY - FROM PUBMETH
BRCA1	12+17	29	22
RASSF1	11+20	31	72
APC	11+15	26	54
CDH1	10+9	19	57
CDKN2A	10+19	29	41
ESR2	7+0	7	58
CCND2	7+5	12	54
ESR1	6+7	13	57
SFRP1	5+2	7	54
RARB	5+15	20	28
GSTP1	4+8	12	31
CDH13	4+1	5	70
SYK	3+0	3	71
MGMT	3+2	5	11
FHIT	3+2	5	41
RUNX3	3+2	5	35
TWIST1	3+7	10	65
TIMP3	3+1	4	51
DAPK1	3+3	6	55
TERT	3+0	3	55
DLC1	3+0	3	17
ID4	3+2	5	38
PYCARD	3+0	3	30
CST6	3+3	6	60
MLH1	2+3	5	25
PLAU	2+0	2	0
HIC1	2+1	3	56
SOCS1	2+0	2	58
SLIT2	2+0	2	49
SCGB3A1	2+1	3	40

PTGS2	2+0	2	64
DSC3	2+0	2	38
DAB2	2+0	2	22
ABCB1	2+1	3	100
KLK10	2+1	3	45
RBP1	2+0	2	19
CADM1	2+0	2	48
MYOD1	2+0	2	85
CDH4	1+0	1	0
TNFRSF12	1+0	1	100
CDH3	1+0	1	58
LAMA3	1+0	1	44
WIF1	1+1	2	67
IGFBP3	1+0	1	18
SLIT3	1+0	1	0
TYMS	1+0	1	5
PAX5	1+0	1	70
LAMB3	1+0	1	4
CLCA2	1+0	1	100
COL1A2	1+0	1	0
SFN	1+0	1	87
SNCG	1+0	1	0
DIRAS3	1+0	1	4
LATS1	1+0	1	57
BCL2	1+0	1	18
ROBO1	1+0	1	19
GADD45A	1+0	1	60
TNFRSF10D	1+0	1	0
TSC1	1+0	1	100
TP73	1+0	1	0
CALCA	1+0	1	98
HRAS	1+0	1	100
TP53	1+0	1	12
PTEN	1+0	1	34
TNFRSF10C	1+0	1	0
CEBPD	1+0	1	11
PRDM2	1+0	1	44
CDCP1	1+0	1	52
TUSC3	1+0	1	9
EPB41L3	1+0	1	27
HSD17B4	1+0	1	82

ZMYND10	1+0	1	0
TGFBR2	1+0	1	30
RB1	1+1	2	17
PRKCDBP	1+0	1	60
SERPINB5	1+0	1	0
CYP1B1	1+0	1	38
WRN	1+0	1	17
LATS2	1+0	1	50
EGFR	1+0	1	0
MCJ	1+0	1	99
GJB2	1+0	1	0
GREM1	1+0	1	54
SULT1A1	1+1	2	64
THBS1	1+0	1	60
CAV1	1+0	1	26
RNR1	1+0	1	100
PGR	1+1	2	100
TMEFF2	1+1	2	35
PLAGL1	1+0	1	0
TRB1	1+0	1	28
TFF1	1+0	1	99
ARHI	1+0	1	99
LAMC2	1+0	1	20
RPRM	1+0	1	0
HS3ST2	1+0	1	88
PCDH10	1+0	1	0
TSC2	1+0	1	17
CDKN1C	1+0	1	45
RARRES1	1+0	1	33

Genes not included in PUBMETH	**REFERENCES from Literature Review only (not included in PUBMETH)**
DLEC1	1
GRIN2B	1
HOXA1	1
MTIG	1
SFRP2	2
SFRP4	1
SFRP5	1

TMEFF2	2
HOXA10	1
IGF2	2
SCGB3A1 (HIN1)	6
DR5	1
DCR1	1
DCR2	1
BRCA2	3
DR4	1
PHD1	1
PHD2	1
PHD3	1
FIH	1
ERα	5
PLCD	1
IL10	1
RIL	1
E-CADHERIN	5
NAT1	2
COMT	1
CYP1A1	1
CYP2D6	1
SYNM	1
Hs.137007	1
Beclin1	1
CIDEA	1
DBC1	1
EMILIN2	1
FBLN2	1
SALL1	1
NRG1	2
XEDAR1	1
CASPASE8	1
MASPIN	1
CXCL	1
THRB	1
FOXC1	2
PPP2R2B	1
PRB	2
STRATIFIN	1
MDR1	2

DAL1	1
CDK10	1
EphA5	1
HSD17B1	1
HSD17B2	1
14-3-3sigma	3
K(v)1.3 potassium channel gene	1
DKK3	2
BIN1	1
BMP6	1
p21	1
SOX17	1
MT1-MMP1/MMP2	1
TIMP2	1
ADAM33	1
HOXB13	2
HNF1B1	1
TLX1	1
ATM	1
EFEMP1	1
Fat4	1
FEN1	1
Hsulf-1	1
ITIH5	1
OPCML	1
CDH5	1
FEZ/LZTS1	1
CXCL12	1
NES1	1
RFC	1
SLC16A12	1
GAL16A12	1
GALR2	1
TOX	1
SPOCK2	1
EGFR5	1
DPYS	1
DFNA5	1
SH3GL2	1

RAD9	1
LDLRAP1	1
GPC3	1
HOXD11	1
LAMA3	1
LAMB3	1
LAMC2	1
ROBO1	1
SYK	1
AK5	1
FOXA2	1
WRN	1
BCL2	1
DCC	1
TMS1	1
P3H2	1
P3H3	1
BMRS1	1
synuclein-γ	2
urokinase	1
N-cadherin	1
ID4	1
annexin A4	1
b-catenin	1
vvNT11	1
B4GALT1	1
KIF1A	1
NISCH	1
Nm23-H1	1
NOEY2	1
OGDHL	1
PAK3	1
RIZ1	1
SERDINB5	1
SOCS1	1
TGFβ RII	1
HMSH2	1
WT-1	1
Gelsolin	1
LDT1	1

MCT1	1
GNA11	1
NORE-1	1
MAGE	1

TOTAL = 231 genes

APPENDIX D

Table 1 – Diet and DNA Methylation Review: Studies on diet and gene specific methylation

Folate					
Author (year)	**Exposure**	**Tissue - Subject counts**	**Control**	**Platform / technique for methylation quantification**	**Main Findings**
Xu (2011)	Folate	Human - Breast cancer tumor tissues (n=800)	Not applicable	Methylation-specific PCR	CCND2 decreased likelihood of methylation with increased intake (P=0.01)
Vineis (2011)	Plasma folate level	Human - Blood (n=186)	Not applicable	Bisulfite Pyrosequencing	RASSF1A, MTHFR: Increased and MGMT, CDKN2A/p16, GSTP1: No change
Kim (2011)	Plasma folate levels - high	Human - Circulating DNA from plasma (n=120)	Medium and Low Plasma Folate	Methylation Specific PCR	p73 - More frequently methylated in the high folate group [50% (8 of 16)] than in the medium [16.7% (3 of 18)] or low folate subgroups [11.1% (2 of 18)]
Kim (2011)	Plasma folate levels - high	Human - Circulating DNA from plasma (n=120)	Medium and Low Plasma Folate	Methylation Specific PCR	p16INK4A - No change (p>0.05)
Kim (2011)	Plasma folate levels - high	Human - Circulating DNA from plasma (n=120)	Medium and Low Plasma Folate	Methylation Specific PCR	hMLH1 - No change (p>0.05)
Wallace (2010)	Folate intake levels	Human - Colon and Rectal biopsies (n=1000)	Not applicable	Bisulfite Pyrosequencing	Oestrogen receptor alpha (*ERα*) - no change in methylation (P trend between quartiles= 0.40)
Wallace (2010)	Folate intake levels	Human - Colon and Rectal biopsies (n=1000)	Not applicable	Bisulfite Pyrosequencing	Secreted frizzled related protein-1 (SFRP1) - no change in methylation (P trend between quartiles= 0.58)
Wallace (2010)	RBC folate levels	Human - Colon and Rectal biopsies (n=1000)	Not applicable	Bisulfite Pyrosequencing	Oestrogen receptor alpha (*ERα*) - increase in methylation (P trend between quartiles= 0.03)
Wallace (2010)	RBC folate levels	Human - Colon and Rectal biopsies (n=1000)	Not applicable	Bisulfite Pyrosequencing	Secreted frizzled related protein-1 (SFRP1) - increase in methylation (P trend between quartiles= 0.01)
Van Guelpen (2010)	Plasma Folate - lowest quantile	Human - Colorectal cancer tumours (n=103)	Plasma Folate - quantiles 2-5	Methyl Light	OR for cancer with low CIMP=2.96 (95% CI 1.24-7.08) for plasma folate quintiles 2-5 versus 1
de Vogel (2011)	Folate intake	Human - Colorectal tumours (n=609)	Not applicable	Methylation Specific PCR	Promoter hypermethylation of at least 3 out of 5 methylation markers (CACNA1G, IGF2, NEUROG1, RUNX3, and SOCS1): no change (P trend between tertiles: 0.54)
Hirsch (2008)	High serum folate	Human - Blood of healthy subjects (n=111)	Low serum folate	Bisulfide treatment and PCR sequencing	Extracellular superoxide dismutase (ec-SOD) gene - No significant change (P-value not given)
Mokarram (2008)	High Folate/VitaminB12 levels	Human - Colorectal cancer tumours (n=151)	Low Folate/VitaminB12 levels	Methylation Specific PCR	p16, hMLH1, and hMSH2 - Odds Ratio for at least one methylated promoter: High folate/Vitamin B12 group: 4.9 (95% CI, 1.4-17.7) Low folate/Vitamin B12 group: 3.9 (95%

					CI, 1.1-13.9)
de Vogel (2008)	Folate	Human - Colorectal cancer tumours (n=734)	Not applicable	Bisulfide treatment and Methylation Specific PCR	MLH1 - No significant change (P-trend between tertiles of intake: men = 0.83, women = 0.76)
Al-Ghnaniem (2007)	RBC Folate	Human - Colorectal mucosa disease free (n=73)	Not applicable	Methyl acceptance assay	ERα - No change and MLH1 - No change (P-values not given)
van den Donk (2007)	Folic Acid and VitaminB12 supplementation	Human - Rectal biopsies in subjects with colorectal adenoma (n=86)	No supplementation	Bisulfide treatment and Methylation Specific PCR	O6-MGMT, hMLH1, p14ARF, p16INK4A, RASSF1A and APC - No significant change (P-values not given)
Mas (2007)	Low levels of folate	Human - Colorectal cancer tumours (n=120)	Cancer free controls (n=296)	Bisulfide treatment and Methylation Specific PCR	p16INK4A - More likely to be methylated in subjects consuming low levels of folate (p=0.01)
Slattery (2007)	Dietary folate	Human - Colon cancer tumour biopsies (n=1154)	Not applicable	Methylation Specific PCR	High CIMP: at least 2 of the following markers methylated: MINT1, MINT2, MINT31, p16 and hMLH1 - Odds ratio for high CIMP in Tertile 3 vs. Tertile 1 of intake: 0.8 (95% CI: 0.6, 1.2)
Kraunz (2006)	Low Dietary folate	Human - Head and neck squamous cell carcinoma biopsies (n=169)	High Dietary folate	Methylation Specific PCR	p16(INK4A) - OR for methylation in the low vs. high intake groups: 2.3 (95% CI = 1.1-4.8)
Kawakami (2003)	Folate intermediates in tissue (5,10-methylenetetrahydrofolate)	Human - Colorectal cancer tumours (n=103)	Not applicable	PCR MethyLight assay	hMLH1, TIMP3, and ARF - Methylation increased (P-values: 0.005, 0.024, 0.022 respectively)
Kawakami (2003)	Folate intermediates in tissue (5,10-methylenetetrahydrofolate)	Human - Colorectal cancer tumours (n=103)	Not applicable	PCR MethyLight assay	p16, MINT2, DAPK, APC - No significant change in methylation (P-values: 0.12, 0.069, 0.79, 0.13 respectively)
Kawakami (2003)	Folate intermediates in tissue (tetrahydrofolate (FH4))	Human - Colorectal cancer tumours (n=103)	Not applicable	PCR MethyLight assay	hMLH1, TIMP3, and ARF - Methylation increased (P-values: 0.023, 0.028, 0.033 respectively)
Kawakami (2003)	Folate intermediates in tissue (tetrahydrofolate (FH4))	Human - Colorectal cancer tumours (n=103)	Not applicable	PCR MethyLight assay	p16, MINT2, DAPK, APC - No significant change in methylation (P-values: 0.18, 0.12, 0.85, 0.20 respectively)
van Engeland (2003)	Low folate/high alcohol	Human - Colorectal cancer tumours sporadic (n=122)	High folate/low alcohol	Methylation Specific PCR	APC-1A, p14ARF, p16INK4A, hMLH1, O6-MGMT, and RASSF1A - no significant change in frequency of hypermethylation (P-values: 0.193, 0.245, 0.891, 0.841, 0.236, 0.172 respectively)
Fang (1997)	Plasma folic acid	Human - Gastric tumour biopsies (n=21)		3H-S-adenosylmethionine (3H-SAM) in the presence of a methylase followed by Southern blotting	c-myc and c-Ha-ras methylation: positive association with plasma folic acid (P<0.05)

Vitamins					
Rawson (2012)	Dietary vitamin D intake	Human - Colorectal tumour biopsies (n=992)	Not applicable	Not available	DKK1: decreased methylation with increased intake (P=0.001)
McKay (2012)	Infant Vitamin B12	Human - blood from infants (n=430)	Not applicable	Pyrosequencing	IGFBP3: inverse correlation with infant vitamin B12 (rho = -0.16, p = 0.007)
Xu (2011)	Vitamin B2 intake	Human - Breast cancer tumor tissues (n=800)	Not applicable	Methylation-specific PCR	CHD1: decreased frequency of methylation with increased intake (p=0.003), TWIST1 and HIN1: increased frequency of methylation with increased intake (p=0.04 and p=0.02 respectively)
Xu (2011)	Vitamin B6 intake	Human - Breast cancer tumor tissues (n=800)	Not applicable	Methylation-specific PCR	HIN1: increased frequency of methylation with increased intake (p=0.04), CDH1 and CCND2: decreased frequency of methylation with increased intake (p=0.05 and p=0.05 respectively)
Xu (2011)	Vitamin B12 intake	Human - Breast cancer tumor tissues (n=800)	Not applicable	Methylation-specific PCR	CDH1: decreased methylation with increased intake (P=0.002)
de Vogel (2011)	Vitamin B2 intake	Human - Colorectal tumours (n=609)	Not applicable	Methylation Specific PCR	Promoter hypermethylation of at least 3 out of 5 methylation markers (CACNA1G, IGF2, NEUROG1, RUNX3, and SOCS1): no change (P trend between tertiles: 0.62)
de Vogel (2011)	Vitamin B6 intake	Human - Colorectal tumours (n=609)	Not applicable	Methylation Specific PCR	Promoter hypermethylation of at least 3 out of 5 methylation markers (CACNA1G, IGF2, NEUROG1, RUNX3, and SOCS1): no change (P trend between tertiles: 0.72)
de Vogel (2008)	Vitamin B2 intake	Human - Colorectal cancer tumours (n=734)	Not applicable	Bisulfide treatment and Methylation Specific PCR	MLH1 - No significant change (p-trend between tertiles of intake: men = 0.92, women = 0.78)
de Vogel (2008)	Vitamin B6	Human - Colorectal cancer tumours (n=734)	Not applicable	Bisulfide treatment and Methylation Specific PCR	MLH1 - Significant increase in men (p-trend between tertiles of intake: men = 0.03, women = 0.27)
Mokarram (2008)	High Folate/VitaminB12 levels	Human - Colorectal cancer tumours (n=151)	Low Folate/VitaminB12 levels	Methylation Specific PCR	p16, hMLH1, and hMSH2 - Odds Ratio for at least one methylated promoter: High folate/Vitamin B12 group: 4.9 (95% CI, 1.4-17.7) Low folate/Vitamin B12 group: 3.9 (95% CI, 1.1-13.9)
van den Donk (2007)	Folic Acid and VitaminB12 supplementation	Human - Rectal biopsies in subjects with colorectal adenoma (n=86)	No supplementation	Bisulfide treatment and Methylation Specific PCR	O6-MGMT, hMLH1, p14ARF, p16INK4A, RASSF1A and APC - No significant change (P-values not given)
Mas (2007)	Low levels of vitamin A	Human - Colorectal cancer tumours (n=120)	Cancer free controls (n=296)	Bisulfide treatment and Methylation Specific PCR	p16INK4A - More likely to be methylated in subjects consuming low levels of vitamin A (p = 0.01)
Mas (2007)	Low levels of vitamin B12	Human - Colorectal cancer tumours (n=120)	Cancer free controls (n=296)	Bisulfide treatment and Methylation Specific PCR	p16INK4A - More likely to be methylated in subjects consuming low levels of vitamin B12 (p = 0.007)
Mas (2007)	Low levels of vitamin A	Human - Colorectal cancer tumours (n=120)	Cancer free controls (n=296)	Bisulfide treatment and Methylation Specific PCR	p4(ARF) - More likely to be methylated in subjects consuming low levels ofvitamin A (p = 0.001)

Mas (2007)	Low levels of vitamin A	Human - Colorectal cancer tumours (n=120)	Cancer free controls (n=296)	Bisulfide treatment and Methylation Specific PCR	hMLH1 - More likely to be methylated in subjects consuming low levels ofvitamin A (p = 0.05)
Slattery (2007)	Dietary vitamin B6	Human - Colon cancer tumour biopsies (n=1154)	Not applicable	Methylation Specific PCR	High CIMP: at least 2 of the following markers methylated: MINT1, MINT2, MINT31, p16 and hMLH1 - Odds ratio for high CIMP in Tertile 3 vs. Tertile 1 of intake: 0.9 (95% CI: 0.6, 1.3)
Slattery (2007)	Dietary vitamin B12	Human - Colon cancer tumour biopsies (n=1154)	Not applicable	Methylation Specific PCR	High CIMP: at least 2 of the following markers methylated: MINT1, MINT2, MINT31, p16 and hMLH1 - Odds ratio for high CIMP in Tertile 3 vs. Tertile 1 of intake: 1.0 (95% CI: 0.7, 1.4)
Methionine					
Xu (2011)	Dietary methionine intake	Human - Breast cancer tumor tissues (n=800)	Not applicable	Methylation-specific PCR	ESR1, PGR, BRCA1, APC, p16, RARβ2, TWIST1, GSTP1, RASSF1A, HIN1, DAPK, CCND2 and CHD1: no association with frequency of methylation (P-values not shown)
Vineis (2011)	Plasma methionine level	Human - Blood (n=186)	Not applicable	Bisulfite Pyrosequencing	RASSF1A: Decreased and MTHFR, MGMT, CDKN2A/p16, GSTP1: No change
de Vogel (2011)	Dietary methionine intake	Human - Colorectal tumours (n=609)	Not applicable	Methylation Specific PCR	Promoter hypermethylation of at least 3 out of 5 methylation markers (CACNA1G, IGF2, NEUROG1, RUNX3, and SOCS1): no change (P trend between tertiles: 0.42)
de Vogel (2008)	Dietary methionine intake	Human - Colorectal cancer tumours (n=734)	Not applicable	Bisulfide treatment and Methylation Specific PCR	MLH1 - No significant change (p-trend between tertiles of intake: men = 0.12, women = 0.83)
Slattery (2007)	Dietary methionine intake	Human - Colon cancer tumour biopsies (n=1154)	Not applicable	Methylation Specific PCR	High CIMP: at least 2 of the following markers methylated: MINT1, MINT2, MINT31, p16 and hMLH1 - Odds ratio for high CIMP in Tertile 3 vs. Tertile 1 of intake: 0.7 (95% CI: 0.5, 1.0)
Alcohol					
Xu (2011)	Alcohol intake	Human - Breast cancer tumor tissues (n=800)	Not applicable	Methylation-specific PCR	ESR1: decreased frequency of methylation with increased intake (P<0.05)
de Vogel (2008)	Alcohol intake	Human - Colorectal cancer tumours (n=734)	Not applicable	Bisulfide treatment and Methylation Specific PCR	MLH1 - No significant change (p-trend between tertiles of intake: men = 0.36, women = 0.29)
Slattery (2007)	Alcohol intake	Human - Colon cancer tumour biopsies (n=1154)	Not applicable	Methylation Specific PCR	High CIMP: at least 2 of the following markers methylated: MINT1, MINT2, MINT31, p16 and hMLH1 - Odds ratio for high CIMP in Tertile 3 vs. Tertile 1 of intake: 1.3 (95% CI: 0.7, 2.2)
Nan (2005)	Alcohol - high consumption	Human - Gastric tumour biopsies (n=110)	Low consumption of alcohol	Methylation Specific PCR	hMLH1 - increased risk for tumours with hypermethylation OR=3.94 (95% CI: 1.21-2.80)

Other dietary exposures					
Milagro (2012)	Monounsaturated fatty acid intake	Human - Blood normal, overweight and obese women undergoing a 16 week weight reduction program (n=60)	Not applicable	Sequenom's MassARRAY	CLOCK gene CpGs 1 & 8 decreased methylation (P<0.01)
Milagro (2012)	Polyunsaturated fatty acid intake	Human - Blood normal, overweight and obese women undergoing a 16 week weight reduction program (n=60)	Not applicable	Sequenom's MassARRAY	CLOCK gene CpGs 1 & 8 increased methylation (P<0.05)
Xu (2011)	Betaine intake	Human - Breast cancer tumor tissues (n=800)	Not applicable	Methylation-specific PCR	APC increased likelihood of methylation with increased intake (P=0.01)
Xu (2011)	Dietary choline intake	Human - Breast cancer tumor tissues (n=800)	Not applicable	Methylation-specific PCR	ESR1, PGR, BRCA1, APC, p16, RARβ2, TWIST1, GSTP1, RASSF1A, HIN1, DAPK, CCND2 and CHD1: no association with frequency of methylation (P-values not shown)
van Breda (2009)	Vegetable intake changes	Human - Colonic mucosa from adenoma patients and healthy controls	Not applicable	Not available	Protein kinase C b 1, ornithine decarboxylase 1, fos proto-oncogene and 5,10-methylenetetrahydrofolate reductase - No Change (P-value not available)
Qin (2009)	Soy isoflavones - >600ng/ml circulating levels of genistein after treatment	Human - Healthy Breast tissue (n=34)	Before treatment levels	Bisulfide treatment and Methylation Specific PCR	RAR-β2: Increased methylation with increased intake (r=0.51, P=0.03)
Qin (2009)	Soy isoflavones - >200ng/ml circulating levels of genistein after treatment	Human - Healthy Breast tissue (n=34)	Before treatment levels	Bisulfide treatment and Methylation Specific PCR	CCND2 - No significant change (r=0.43, P=0.29)
Qin (2009)	Soy isoflavones (any post-treatment concentration)	Human - Healthy Breast tissue (n=34)	Before treatment levels	Bisulfide treatment and Methylation Specific PCR	p16, RASSF-1A, ER - no significant change (P-values not shown)
Thaler (2009)	Vegetarian diet	Human - Buccal mucosa (n=120)	Omnivore diet - matched age	Methylation Specific PCR	Antioxidative enzyme mitochondrial superoxide dismutase (MnSOD) - Decreased methylation (P-value not available)
Premkumar (2008)	Daily supplement of 100 mg co-enzyme Q10, 10 mg riboflavin and 50 mg niacin	Human - Blood from Breast Cancer patients (n=84)	Whole blood from patients without supplementation	Bisulfide treatment and Methylation Specific PCR	RASSF1A - disappearance of methylation in the treated group (P-value not given)
Mas (2007)	Low levels of potassium	Human - Colorectal cancer tumours (n=120)	Cancer free controls (n=296)	Bisulfide treatment and Methylation Specific PCR	p16INK4A - More likely to be methylated in subjects consuming low levels of potassium (p = 0.03)
Mas (2007)	Low levels of iron	Human - Colorectal cancer tumours (n=120)	Cancer free controls (n=296)	Bisulfide treatment and Methylation Specific PCR	p16INK4A - More likely to be methylated in subjects consuming low levels of iron (p = 0.02)

Nan (2005)	Potatoes - high consumption	Human - Gastric tumour biopsies (n=110)	Low consumption of potatoes	Methylation Specific PCR	hMLH1 - decreased risk for tumours with hypermethylation OR: 0.30 (95%CI: 0.14-0.67)
Nan (2005)	Butter cheese and margerine - high consumption	Human - Gastric tumour biopsies (n=110)	Low consumption of butter cheese and margerine	Methylation Specific PCR	hMLH1 - decreased risk for tumours with hypermethylation OR: 0.44 (95%CI: 0.20-0.93)
Nan (2005)	Vegetables - high consumption	Human - Gastric tumour biopsies (n=110)	Low consumption of vegetables	Methylation Specific PCR	hMLH1 - increased risk for tumours with hypermethylation OR=2.17 (95% CI: 1.03-4.58)
Yuasa (2004)	Green tea - >=7cups/day	Human - Gastric carcinomas primary (n=73)	Green tea - <=6cups/day	Methylation Specific PCR	CDX2 - Decreased methylation (β=-0.37, P=0.02)
Yuasa (2004)	Full or overeating habits	Human - Gastric carcinomas primary (n=73)	Light/moderate eating	Methylation Specific PCR	CDX2 - Increased methylation (β=0.31, P=0.02)

Table 2 – Diet and DNA Methylation Review: Studies on diet and global methylation

Folate

Author (year)	Exposure	Tissue - Subject counts	Control (if application)	Platform / technique for global methylation quantification	Main Findings
Ono (2013)	Folate intake	Human - Blood of Japanese women (n= 384)	Not applicable	LUMA	0.36% decrease in global methylation per quartile increase in folate intake
Perng (2012)	RBC folate	Human - Blood from children 5-12 years (n=568)	Not applicable	Pyrosequencing - LINE1 repetitive elements	No association
Wang (2012)	Serum folate deficiency	Human - Peripheral Blood of Chromate workers (n=115)	60 local residents	ELISA kit	Global hypomethylation (1.48%, as compared to 2.10% observed in the control group, ($p < 0.001$))
Crider (2012)	Folic acid supplementation (6months) of women in a double blind RCT	Human - coagulated blood (n=1108)	Women on placebo	LC-MS/MS analysis of 5-methyl-deoxycytidine	No change in global methylation after supplementation or after withdrawal
Crider (2012)	Folic acid supplementation (6months) of women in a double blind RCT	Human - coagulated blood (n=1108)	Women on placebo	LC-MS/MS analysis of 5-methyl-deoxycytidine	No change in global methylation after supplementation
Crider (2012)	1month after withdrawal of supplementation	Human - coagulated blood (n=1108)	Not applicable	LC-MS/MS analysis of 5-methyl-deoxycytidine	14% decrease in global methylation
Crider (2012)	3months after withdrawal of supplementation	Human - coagulated blood (n=1108)	Not applicable	LC-MS/MS analysis of 5-methyl-deoxycytidine	23% decrease in global methylation
Crider (2012)	3months after withdrawal of supplementation	Human - coagulated blood w/ MTHFR polymorphism (n=363)	Not applicable	LC-MS/MS analysis of 5-methyl-deoxycytidine	>=25% decrease in global methylation
Jung (2011)	Folic Acid supplementation for 3 years	Human - Blood of women with moderately increased homocysteine (n=105)	111 women on placebo treatment	Liquid chromatography-tandem mass spectrometry	No change (difference between treatment groups=0.008, 95%CI=−0.05,0.07, P=0.79)
Elworthy (2011)	A thirteen week folic acid avoidance period and a subsequent eleven week folic acid supplementation period (400 µg folic acid/day)	Human - Blood from healthy individuals (n= 93)	Not applicable	Methylation-sensitive quantitative PCR of Sat-2 regions	Negligible change (Difference after repletion following avoidance: -0.01 ± 0.44(SD) *not statistically significant
Christensen (2010)	Increased Folate Intake	Human - Breast tumour biopsies (n=162)	Not applicable	Illumina GoldenGate methylation bead-array	Trend towards increased methylation (P=0.08)

				platform	
Schernhammer (2010)	Folate (> or = 400mg/day)	Human - Colon cancer tumours (n=609)	<200mg of folate	Pyrosequencing technology using the PyroMark kit and the PSQ HS 96 System	Less likely to develop tumours with LINE-1 hypomethylation
Figueiredo (2009)	Folate dietary intake in quartiles	Human - normal colonic mucosa biopsies (n=388)	Not applicable	Pyrosequencing assay for LINE-1	Increased methylation with increased intake (P for trend: 0.02)
Figueiredo (2009)	Plasma folate in quartiles	Human - normal colonic mucosa biopsies (n=388)	Not applicable	Pyrosequencing assay for LINE-1	No change in methylation with increasing quartile (P for trend: 0.22)
Figueiredo (2009)	RBC folate in quartiles	Human - normal colonic mucosa biopsies (n=388)	Not applicable	Pyrosequencing assay for LINE-1	No change in methylation with increasing quartile (P for trend: 0.99)
Pilsner (2007)	Plasma folate - high (n=137)	Human - Blood DNA of Bagladeshi adults (n=294)	Plasma folate - low (n=87)	Methyl acceptance assay	Increased methylation in the high plasma folate group (P-value: 0.03)
Piyathilake (2006)	Post folic acid fortification of grain products (n=172)	Human - Cervical intraepithelial neoplasia biopsies	Pre- folic acid fortification (n=152)	Monoclonal antibody specific for 5 methyl cytosine (5-mc)	No difference in methylation (P>0.05)
Basten (2006)	Folate supplementation group (n=30)	Human - Blood of healthy volunteers (n=61)	Placebo treatment group (n=31)	Incorporation of methyl groups from 3H-labeled S-adenosyl-L-methionine at specific cytosine residues using the bacterial enzyme Sss1 methylase	Non significant difference - values not given)
Pufulete (2005)	Folic Acid supplementation	Human - Blood and colonic mucosa biopsies of colorectal adenoma patients (n= 10)	Placebo treatment	In vitro [3H] methyl acceptance assay.	Increased methylation relative to placebo - 31% (95% CI: 16-47%) in WBC and 25% (95% CI: 11-39%) in colonic mucosa
Pufulete (2005)	Folate status (RBC and serum)	Human - Colonic mucosa biopsies of healthy subjects (n=68)		In vitro radiolabelled methyl acceptance assay	Folate inversely associated with hypomethylation (Serum folate r=-0.311, P=0.01 and RBC folate r=-0.356, P=0.03 but P=0.07 and P=0.08, respectively when adjusted)
Friso (2002)	Increased RBC Folate	Human - Blood DNA (n=292): subjects homozygous for this mutation (T/T) and subjects homozygous for the wild-type (C/C) MTHFR genotype		Liquid chromatography/MS method	Increased methylation (Individuals with TT: High folate mean methylation: 67.56(50.45–90.38) Low folate mean methylation: 26.21(18.81–36.49) difference: P<0.0001)
Rampersaud (2000)	Folate depletion (moderate)	Human - Blood of elderly women (n=33)		Ability of DNA to incorporate [(3)H]methyl groups from labeled S:-adenosylmethionine in an in vitro assay	Decreased methylation following depletion (ANCOVA P value=0.0025) - exact means not given
Rampersaud (2000)	Folate repletion (after moderate depletion)	Human - Blood of elderly women (n=33)		Ability of DNA to incorporate [(3)H]methyl groups from labeled S:-adenosylmethionine in an in vitro assay	No significant change in methylation following repletion - exact means not given
Piyathilake (2000)	Increasing folate levels	Human - Squamous cell lung cancer tissue biopsies (n=12)		Radiolabeled methyl incorporation	Significantly lower methylation (P value=0.03)
Piyathilake	Increasing folate levels	Human - Normal lung tissue adjacent to squamous cell		Radiolabeled methyl	Increased methylation (P value not available)

(2000)		lung cancer (n=12)		incorporation	
Fowler (1998)	Folate status (Cervical tissue and Serum)	Human - Cervical tissue biopsies (n=83)		[3H]methyl-S-adenosylmethionine and Sss 1 methylase.	Folate significantly correlated with DNA methylation (r=0.2793, P = 0.0211 for cervical tissue and r=2392, P = 0.0569 for serum)
Fenech (1998)	RBC Folate	Human - Blood (n=106)		[3H]methyl-S-adenosylmethionine and Sss 1 methylase.	No significant correlations with DNA methylation (Males: -0.0160 and Females: 0.0893)
Fenech (1998)	Folate and vitamin B12 (B12) supplementation	Human - Blood (n=33)	Before supplementation values	[3H]methyl-S-adenosylmethionine and Sss 1 methylase.	No significant change in methylation (Umethylated CpGs at R1, R2 and R3: 195200, 204500, and 203800 respectively)
Jacob (1998)	Dietary folate	Human - Blood from menopausal women (n=8)		[3H-methyl] S-adenosyl L-methionine and Sss I CpG methylase	Inversely associated with hypomethylation (DNA methyl acceptance at baseline: 795+/-41, after depletion: 1738+/-259*, after repletion: 1084+/-158*; *significantly different results compared to previous mean)
Cravo (1994)	Folic acid supplementation (10mg/day)	Human - rectal mucosa of patients with resected colonic adenomas (n=20)		[3H]methyl-S-adenosylmethionine and Sss 1 methylase.	Decreased hypomethylation methylation in 7 out of the 20 subjects (P=0.05) (all with a single polyp)
	Serum folate	Human - rectal mucosa of patients with resected colonic adenomas (n=20)		[3H]methyl-S-adenosylmethionine and Sss 1 methylase.	Inverse correlation between serum folate and DNA methylation (r = -0.56; P = 0.01).

Alcohol

First Author (year)	Exposure of Interest (own or maternal)	Cell type	Control	Platform/Technique	Global Methylation Level
Ono (2013)	Alcohol, Vitamins B2 B6 and B12	Human - Blood of Japanese women (n= 384)	Not applicable	LUMA	No association - but alcohol modifies the association between folate and methylation
Christensen (2010)	Increased Alcohol Intake	Human - Breast tumour biopsies (n=162)	Not applicable	Illumina GoldenGate methylation bead-array platform	Trend towards decreased methylation (P=0.07)
Schernhammer (2010)	Alcohol (> or = 15g/day)	Human - Colon cancer tumours (n=609)	<200mg of folate	Pyrosequencing technology using the PyroMark kit and the PSQ HS 96 System	More likely to develop tumours with LINE-1 hypomethylation (RR=1.67, 95% CI=1.04 to 2.67 for <55% LINE1 methylated tumours; RR=1.55, 95% CI=1.10 to 2.18 for 55-64% LINE-1 methylated tumours; and RR=1.06, 95% CI=0.69 to 1.62 for > or = to 65% LINE1 methylated tumours)
Figueiredo (2009)	Alcohol (intake in drinks per day)	Human - normal colonic mucosa biopsies (n=388)	Not applicable	Pyrosequencing assay for LINE-1	No change across categories of intake (P for trend: 0.74)
Bonsch (2006)	Chronic alcoholism (59 subjects)	Human - Blood (n=125)	No alcoholism (66 subjects)	Modified non-radioactive elongationassay	10% increased methylation (Chronic alcoholism: 61% SD: 16.7, Healthy: 51% SD:20.2; P-value=0.03)

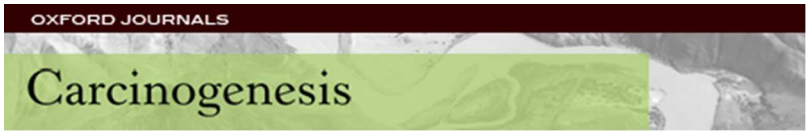
B Vitamins					
First Author (year)	**Exposure of Interest (own or maternal)**	**Cell type**	**Control**	**Platform/Technique**	**Global Methylation Level**
Perng (2012)	Plasma Vitamin B12	Human - Blood from children 5-12 years (n=568)	Not applicable	Pyrosequencing - LINE1 repetitive elements	No association
Ono (2013)	Alcohol, Vitamins B2 B6 and B12	Human - Blood of Japanese women (n= 384)	Not applicable	LUMA	No association - but lacohol modifies the association between folate and methylation
Schernhammer (2010)	B6 vitamins (5th quartile of intake)	Human - Colon cancer tumour biopsies (n=606)	(1st quartile - lowest intake)	PyroMark kit and the PSQ HS 96 System (Biotage, Uppsala, Sweden)	No effect on likelihood of developing tumours with LINE-1 hypomethylation ((RR=0.84, 95% CI=0.52 to 1.36 for <55% LINE1 methylated tumours; RR=0.71, 95% CI=0.48 to 1.05 for 55-64% LINE-1 methylated tumours; and RR=1.20, 95% CI=0.77 to 1.86 for > or = to 65% LINE1 methylated tumours)
	B12 vitamins (5th quartile of intake)	Human - Colon cancer tumour biopsies (n=606)	(1st quartile - lowest intake)	PyroMark kit and the PSQ HS 96 System (Biotage, Uppsala, Sweden)	No effect on likelihood of developing tumours with LINE-1 hypomethylation (RR=0.62, 95% CI=0.37 to 1.06 for <55% LINE1 methylated tumours; RR=0.89, 95% CI=0.61 to 1.30 for 55-64% LINE-1 methylated tumours; and RR=0.73, 95% CI=0.47 to 1.13 for > or = to 65% LINE1 methylated tumours)
Figueiredo (2009)	B2 dietary intake in quartiles	Human - normal colonic mucosa biopsies (n=388)	Not applicable	Pyrosequencing assay for LINE-1	No chnage with increased intake (P for trend: 0.57)
	B2 plasma levels in quartiles	Human - normal colonic mucosa biopsies (n=388)	Not applicable	Pyrosequencing assay for LINE-1	No change with increasing quartile (P for trend: 0.45)
	B6 dietary intake in quartiles	Human - normal colonic mucosa biopsies (n=388)	Not applicable	Pyrosequencing assay for LINE-1	No change with increased intake (P for trend: 0.17)
	B6 plasma levels in quartiles	Human - normal colonic mucosa biopsies (n=388)	Not applicable	Pyrosequencing assay for LINE-1	No change with increasing quartile (P for trend: 0.29)
	B12 plasma levels in quartiles	Human - normal colonic mucosa biopsies (n=388)	Not applicable	Pyrosequencing assay for LINE-1	No change with increasing quartile (P for trend: 0.58)
Piyathilake (2000)	Increasing B12 levels	Human - Squamous cell lung cancer tissue biopsies (n=12)	Not applicable	Radiolabeled methyl incorporation	Significantly lower methylation (P value=0.009)
		Human - Normal lung tissue adjacent to squamous cell lung cancer (n=12)		Radiolabeled methyl incorporation	No significant relationship
Fenech (1998)	Plasma vitamin B12	Human - Blood (n=106)	Not applicable	[3H]S-adenosyl methionine and Sss1 methylase	Significant correlation with DNA methylation in women (Males: -0.0557 and Females: -0.3593*)
	Folate and vitamin B12	Human - Blood (n=33)	Before	[3H]S-adenosyl methionine	No significant change in methylation (Umethylated CpGs at

	(B12) supplementation		supplementation values	and Sss1 methylase	R1, R2 and R3: 195200, 204500, and 203800 respectively)

Other Dietary Exposures					
First Author (year)	**Exposure of Interest (own or maternal)**	**Cell type**	**Control**	**Platform/Technique**	**Global Methylation Level**
Zhang (2011)	Prudent dietary pattern (Highest intake Quartile) - characterized by a high intake of vegetables and fruits	Human - Blood (n=149	Lowest quartile of intake	Bisulfite conversion of DNA and real-time PCR (MethyLight) for LINE-1	Decreased prevalence of hypomethylation (Q4 vs. Q1: OR=0.33, 95% CI:0.12-0.92)
Perng (2012)	Plasma Vitamin A - levels higher than or equal to 1.05 µmol/L vs. 0.70 or lower	Human - Blood from children 5-12 years (n=568)	Not applicable	Pyrosequencing - LINE1 repetitive elements	0.19% decrease in global methylation
Wang 2012	Serum folate deficiency in Chromate workers (115)	Human - Peripheral Blood of Chromate workers (n=115)	60 local residents	ELISA kit	Global hypomethylation (1.48%, as compared to 2.10% observed in the control group, ($p < 0.001$))
Perng (2012)	Ferritin	Human - Blood from children 5-12 years (n=568)	Not applicable	Pyrosequencing - LINE1 repetitive elements	No association
Perng (2012)	Serum zinc	Human - Blood from children 5-12 years (n=568)	Not applicable	Pyrosequencing - LINE1 repetitive elements	No association

APPENDIX E

Towards incorporating epigenetic mechanisms into carcinogen identification and evaluation



Towards incorporating epigenetic mechanisms into carcinogen identification and evaluation

Journal:	*Carcinogenesis*
Manuscript ID:	CARCIN-2013-00292.R2
Manuscript Type:	Review
Date Submitted by the Author:	31-May-2013
Complete List of Authors:	Zdenko, Herceg; IARC, MCA/EGE Lambert, Marie Pierre; IARC, MCA/EGE van Veldhoven, Karin; Imperial College London, Epidemiology and Biostatistics Demetriou, Christiana; Imperial College London, Epidemiology and Biostatistics Vineis, Paolo; Imperial College London, Epidemiology and Biostatistics Smith, Martyn; UC Berkeley , School of Public Health ; Straif, Kurt; IARC, Monographs Wild, Christopher; International Agency for Research on Cancer,
Keywords:	epigenetic mechanisms, carcinogen identification and evaluation; IARC Monographs, environment, DNA methylation, histone modifications

SCHOLARONE™
Manuscripts

Downloaded from http://carcin.oxfordjournals.org/ at Imperial College London Library on June 15, 2013

Towards incorporating epigenetic mechanisms into carcinogen identification and evaluation

Zdenko Herceg[1], Marie-Pierre Lambert[1], Karin van Veldhoven[2], Christiana Demetriou[2], Paolo Vineis[2], Martyn T. Smith[3], Kurt Straif[1], Christopher P. Wild[1]

[1]*International Agency for Research on Cancer (IARC), 150 Cours Albert Thomas, 69008, Lyon, France*
[2]*Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK*
[3]*School of Public Health Division of Environmental Health Sciences University of California Berkeley, CA 94720-7356*

Running head: Epigenetic mechanisms in carcinogen evaluation

Corresponding author: Zdenko Herceg, PhD
Epigenetics Group
International Agency for Research on Cancer (IARC)
150 Cours Albert Thomas, F-69008, Lyon, France
Tel: +33 4 72 73 83 98 ; Fax: +33 4 72 73 83 22
E-mail: herceg@iarc.fr

Key words: epigenetic mechanisms, carcinogen identification and evaluation, environment, DNA methylation, histone modifications, IARC Monographs

For Peer Review

Downloaded from http://carcin.oxfordjournals.org/ at Imperial College London Library on June 15, 2013

Abstract

Remarkable progress in the field of epigenetics has turned academic, medical and public attention to the potential applications of these new advances in medicine and various fields of biomedical research. The result is a broader appreciation of epigenetic phenomena in the etiology of common human diseases, most notably cancer. These advances also represent an exciting opportunity to incorporate epigenetics and epigenomics into carcinogen identification and safety assessment. Current epigenetic studies, including major international sequencing projects, are expected to generate information for establishing the “normal” epigenome of tissues and cell types as well as the physiological variability of the epigenome against which carcinogen exposure can be assessed. Recently, epigenetic events have emerged as key mechanisms in cancer development, and while our search of the Monograph Volume 100 revealed that epigenetics have played a modest role in evaluating human carcinogens by the International Agency for Research on Cancer (IARC) Monographs so far, epigenetic data might play a pivotal role in the future. Here, we review (i) the current status of incorporation of epigenetics in carcinogen evaluation in the IARC Monographs Programme, (ii) potential modes of action for epigenetic carcinogens, (iii) current in vivo and in vitro technologies to detect epigenetic carcinogens, (iv) genomic regions and epigenetic modifications and their biological consequences, and (v) critical technological and biological issues in assessment of epigenetic carcinogens. We also discuss the issues related to opportunities and challenges in the application of epigenetic testing in carcinogen identification and evaluation. Although the application of epigenetic assays in carcinogen evaluation is still in its infancy, important data are being generated and valuable scientific resources are being established that should catalyze future applications of epigenetic testing.

For Peer Review

Downloaded from http://carcin.oxfordjournals.org/ at Imperial College London Library on June 15, 2013

Introduction

Epigenetics is a rapidly expanding field of modern biology with a profound impact on our thinking and understanding of biological phenomena. The term “epigenetic” refers to all stable changes in gene expression and chromatin organization that are independent of the DNA sequence itself and that can be mitotically inherited over cell divisions. Epigenetic phenomena, including genomic imprinting, X-chromosome inactivation and global reconfiguration of the DNA methylome, changes in chromatin compaction states and histone modification patterns, occur during development and contribute to the lineage-specific epigenome that is maintained over the lifetime of an organism.

Epigenetic mechanisms are essential for the stable propagation of gene activity states from one generation of cells to the next, and thus the epigenome governs the establishment and maintenance of cell identity. DNA methylation, histone modifications and non-coding RNAs are the main epigenetic mechanisms that may act alone or in combination to govern the gene expression programme over the lifetime of an organism.

The remarkable progress in the field of epigenetics has turned academic and medical attention to the potential application of new conceptual advances in cancer research. The advent of novel technologies that allow cost-effective profiling of the epigenome with unprecedented resolution has dramatically accelerated cancer research and opened up new perspectives. Together, these advances have led to a broader appreciation of epigenetics in the etiology of complex human diseases, including cancer. In addition to their application in mechanistic studies, cancer therapy and biomarker discovery, advances in epigenetics also need to be incorporated into carcinogen identification and safety assessment.

Unlike the genome, which is virtually identical among all cells within an organism, different tissues and cell types harbour a distinct epigenome, which may undergo substantial changes with ageing and in response to environmental factors (1,2). Epigenetic mechanisms can be viewed as an interface between the environment and the genome, the deregulation of which may disrupt key cellular processes, ultimately resulting in oncogenic transformation and tumour development (Figure 1). Exposure to environmental factors may leave a fingerprint on the epigenome that may be exploited in discovering new biomarkers for risk assessment and cancer prevention. Despite an increased interest in epigenetics and a better understanding of epigenetic mechanisms and their deregulation in human malignancies, relatively little is known about the potential application of epigenetics in

For Peer Review

Downloaded from http://carcin.oxfordjournals.org/ at Imperial College London Library on June 15, 2013

carcinogen identification and evaluation. While epigenetic assays have not yet been incorporated into the current carcinogen testing battery, recent evidence on the impact of environmental exposures on the epigenome argues that the time is ripe for established and suspected carcinogens to be specifically evaluated for their potential to deregulate epigenetic mechanisms. The gaps in our understanding of the normal variability of the epigenome in different cell types should be bridged by major international sequencing initiatives enabled by the genome/epigenome technology revolution. Once this is available it will be possible to identify specific changes in the epigenome associated with exposures to a carcinogen. Several recent reviews on the impact of environmental/lifestyle factors on the epigenome have been published (3-7).

The advances in understanding the normal epigenome should also facilitate development of epigenetic assays in carcinogen testing and use of epigenetic data in the regulatory process and policy-making. In this review, we will summarize the current status of incorporation of epigenetics in carcinogen evaluation by the International Agency for Research on Cancer (IARC) Monographs Programme. We will also assess the selected examples of studies demonstrating the epigenetic effects of specific agents in human carcinogenesis. Finally, we will discuss the issues related to opportunities and challenges in incorporating epigenetics into carcinogen evaluation and future perspectives.

Physiological and pathological changes of the epigenome

An important feature of the epigenome is that it is susceptible to normal variations, that is to say that epigenetic modifications of DNA and histones exhibit variations across different cell types and over time. It is thought that the genome needs to be plastic enough to respond to environmental stressors and endogenous cues (4,8). Such induced changes in the epigenome may be fixed and propagated over cell divisions, resulting in long-term or permanent changes in phenotype. Therefore, in contrast to genotype, modulation of the epigenotype is a physiological and essential process that controls an organism's response to environmental exposures.

The reconfiguration of the epigenome is particularly evident during embryonic development, cell differentiation and ageing (9-11). A complete erasure and re-establishment of the DNA methylation patterns in early embryonic development is a dynamic, but highly regulated, process that results in a profound reconfiguration of the

For Peer Review

Downloaded from http://carcin.oxfordjournals.org/ at Imperial College London Library on June 15, 2013

epigenome (including the DNA methylome but also histone marks) (12) and is believed to be highly susceptible to environmental exposures. The reconfiguration of the methylome during early embryonic development provides a striking example of dynamic changes in the epigenome. In the fertilized egg before the fusion of the pronuclei, DNA methylation levels in the maternal and paternal genomes are heavily methylated and both pronuclei have roughly the same methyl-cytosine content (12,13). Soon after the fusion (within approximately an hour) the paternal genome is efficiently demethylated through an active mechanism (active demethylation), and its demethylation levels remain low during a few mitotic divisions. In contrast, the maternal genome undergoes a passive demethylation during several subsequent mitoses. The genome of the early embryo starts to be remethylated by *de novo* DNA methylases at implantation, and the establishment of DNA methylation patterns occurs in a tissue-specific manner. It is believed that this process of dramatic demethylation/remethylation represents a window of susceptibility to environmental stressors and that adverse changes in the epigenome during early embryonic development may be at the heart of intra-uterine programming of childhood and adult diseases (14). Another stage during embryonic development that may be considered vulnerable is the primordial germ cells (PGCs) (15,16). At this stage, foetal germline is dynamically remodelled in a gender-specific manner and epigenetic modifications (such as DNA methylation marking of imprinted genes) are removed (12,17). However, PGCs can contribute genetic material to the future offspring of the foetus, therefore epigenetic changes induced by carcinogen exposure during this stage may contribute to transgenerational epigenetic inheritance. It is therefore essential to account for these events when evaluating carcinogens.

Despite a better understanding of epigenetic mechanisms and their deregulation in human cancer, much remains unknown about the normal dynamic variations of the epigenome and how to distinguish them from adverse epigenetic changes that pose a health risk. In general, the degree to which agents may promote carcinogenesis through epigenetic mechanisms depends on the amount and duration of exposure. The degree to which an adverse exposure alters the epigenome may also strongly depend on variation in susceptibility to the exposure. Individual susceptibility to a given exposure is likely to depend on the epigenetic make-up that dictates an individual's response and adaptation mechanisms. Differences in individual susceptibility may be attributed to patterns of DNA methylation, histone modifications and non-coding RNAs, as well as genetic make-up.

For Peer Review

Downloaded from http://carcin.oxfordjournals.org/ at Imperial College London Library on June 15, 2013

Genotoxic, non-genotoxic, and epigenetic carcinogens

Epigenetic mechanisms are thought to play important roles in the adaptation and response to environmental exposures, although a clear-cut causal relationship between epigenetic change and specific exposure is often difficult to establish. The reason for this may be three-fold. First, environmental agents may induce subtle changes and quantitative phenotypic manifestation may be evident only after repetitive or prolonged exposure. Second, environmental exposures are likely to induce global non-specific changes in one layer or multiple layers of the epigenome. Third, there is an important gap in our understanding of the "normal" epigenome in tissues and cell types and the normal variability of the epigenome.

Recent studies have implicated epigenetic mechanisms in carcinogenesis linked to environmental exposures, although there is a paucity of evidence demonstrating molecular mechanisms by which the epigenome is deregulated in response to a specific carcinogen. A wide range of known and suspected carcinogens (including chemical, physical and biological agents) have been associated with changes in the epigenome, and it has been suggested that their mode of action may involve disruption of epigenetic mechanisms (Table I-IV). The effects of these carcinogens on epigenetic states have been either demonstrated experimentally using different animal and cellular models or inferred from epidemiological studies (4,7). Environmental factors associated with epigenetic deregulation include tobacco smoke, arsenic, cadmium, nickel and ionizing and UV radiation. Different infectious agents such as human hepatitis B virus (HBV), hepatitis C virus (HCV), and human papillomavirus (HPV) and the bacterium *Helicobacter pylori* have been shown to deregulate proliferation, cell division and the gene expression pattern of the host cell via an epigenetic strategy. Among dietary factors, alcohol and fat consumption may act epigenetic carcinogens.

It should be noted that many carcinogens may promote tumour development through the induction of both epigenetic changes (aberrant DNA methylation and histone modifications) and genetic alterations (mutations). However, individual genetic polymorphisms and epigenetic make-up ("epigenetic polymorphisms") may also play pivotal roles in cellular response to environmental stress and thus may represent a part of individual predisposition to the development of cancer. Therefore, individual cancer susceptibility is

For Peer Review

Downloaded from http://carcin.oxfordjournals.org/ at Imperial College London Library on June 15, 2013

likely to depend not only on genetic but also on epigenetic make-up. These responses involve the action of diverse cellular machineries such as those involved in DNA repair, carcinogen detoxification, cell cycle control and cell death. The technological advances in epigenomics (high-throughput and genome-wide profiling) will soon allow the identification of entire epigenomes (genome-wide patterns of DNA methylation and histone modifications). This may provide critical information for testing the notion that differences in individual susceptibility may also be attributed to germline epigenetic make-up.

In the literature, the terms “non-genotoxic” and “epigenetic” are sometimes used interchangeably (18). However, not all non-genotoxic carcinogens act by altering epigenetic states (DNA methylation, histone modifications or non-coding RNAs). For example, some non-genotoxic agents, such as dioxin, can act via receptor-mediated pathways (19), which cannot be considered as an epigenetic mechanism. This means that while some epigenetic agents are also genotoxic, not all non-genotoxic agents are epigenetic carcinogens and indeed, as discussed below, agents may act by multiple mechanisms. Therefore, oversimplified classification of carcinogens that does not consider potential contributing effects that are neither genetic or epigenetic is confusing and should be avoided.

The IARC Monographs Programme

IARC Monographs. The IARC Monographs Programme identifies environmental and lifestyle factors that are human carcinogens (20). Interdisciplinary Working Groups of expert scientists review published studies and evaluate the strength of the evidence that an agent can increase the risk of human cancer. Since 1971, more than 950 agents have been evaluated, of which more than 400 have been identified as carcinogenic, probably carcinogenic, or possibly carcinogenic to humans. Recently, IARC completed a review of the more than 100 chemicals, occupations, physical agents, biological agents, and other agents classified as carcinogenic to humans (21).

Volume 100 of the IARC Monographs reviews all agents previously classified by IARC as carcinogenic to humans (Group 1) and is divided into 6 parts (A,B,C,D,E,F), each of which describes a distinct class of carcinogens. Our search of the Monograph Volume 100 revealed that only a few chemical agents and nutritional or lifestyle factors evaluated by IARC Working Groups have been considered as epigenetic carcinogens. Diethylstilbestrol

For Peer Review

Downloaded from http://carcin.oxfordjournals.org/ at Imperial College London Library on June 15, 2013

(DES), chromium (VI) compounds and ionizing radiation have been classified as carcinogens that may act through epigenomic deregulation. Curiously, for all of these, their impact on the epigenome is considered not as the major mechanism but rather as the secondary mechanism. The lack of focus on epigenetics in relation to the mechanistic data is perhaps surprising considering a wealth of studies demonstrating the impact of many of the Volume 100 carcinogens on epigenetic mechanisms (Tables I-IV). One explanation may lie in the fact that many studies failed to address whether changes in the epigenome were causal or associative to a carcinogenic exposure. In this section, we discuss the specific example of studies demonstrating the epigenetic effects of DES and infectious agents in human carcinogenesis.

DES – an example of an epigenetic carcinogen from the IARC Monographs? DES is a synthetic estrogen that was widely used (from the 1940s to the 1970s in the USA) to prevent potential miscarriages (through its stimulation of placental synthesis of estrogen and progesterone) and for the treatment of symptoms associated with menopause and ovariectomy as well as specific vaginal and vulvar conditions (such as inflammation and dystrophy). DES was also used as a post-coital emergency contraceptive ("morning-after pill") and to treat other conditions associated with dysfunctional menstrual cycles and female hypogonadism. It has been estimated that 5–10 million US citizens were treated with DES during pregnancy or were exposed to the drug *in utero* (22). Although nowadays DES is rarely used (e.g. to treat prostate cancer or specific forms of breast cancer), the consequences of its use are still felt among the treated individuals and their progeny.

DES had been evaluated by IARC Working Groups in 1978 and 1987 and was evaluated again for Volume 100 (in 2008); IARC has classified this chemical as a Group 1 carcinogen (21). The IARC Monograph of 1987 states that there is sufficient evidence of a causal association between clear-cell adenocarcinoma (CCA) of the vagina or cervix and prenatal exposure to DES (23). It also states that there is sufficient evidence of a causal relationship between cancer of the breast and the use of DES during pregnancy. The carcinogenicity of DES was established or highly suspected in experimental animals before epidemiological studies confirmed its carcinogenicity in humans. Consistent with its estrogenic properties, DES was shown to induce various effects on the reproductive system in both mice and humans (22,24,25). In a mouse model, it was shown that prenatal and perinatal exposure to DES produces multiple effects in uterine tissue, including uterine cancer (leiomyomas) (25).

For Peer Review

Downloaded from http://carcin.oxfordjournals.org/ at Imperial College London Library on June 15, 2013

Several studies have investigated the mechanism by which DES exposure promotes carcinogenesis. These include studies of gene expression and DNA methylation states in uterine tissue. Specific changes in gene expression in the uterus of young mice treated neonatally were detected after exposure to DES (24). These alterations were detected in specific genes (fos and lactoferrin) and persisted for weeks even after treatment cessation. Interestingly, gene expression changes were associated with specific epigenetic changes, namely changes in DNA methylation. For example, the genes that were differentially expressed in DES-treated animals also exhibited abnormal DNA methylation (24,26).

The above results strongly suggest that exposure to DES may have a significant and long-term effect on gene expression through epigenetic mechanisms. However, it should be noted that these studies focused on only a few genes and one epigenetic mechanism (DNA methylation). The impact of DES is unlikely to be limited to a small subset of genes, and epigenome-wide studies have yet to be performed. Indeed, microarray-based transcriptome analysis in both rats and mice has revealed DES-induced changes in expression of a wide range of genes, although whether these changes were accompanied by changes in DNA methylation states or other epigenetic alterations (histone modifications and non-coding RNAs) has been little studied (27-29).

Another interesting feature of DES exposure is its impact on cancer incidence in subsequent generations. It has been shown that in addition to an increased cancer susceptibility associated with epigenetic changes in DES-treated parents, an epigenetic mechanism may operate in subsequent generations of mice (the F2 generation) (24). These findings further support the notion that DES-induced carcinogenesis may operate through an epigenetic mechanism, although further studies extending to the F3 generation of exposed animals are needed in order to establish a true transgenerational epigenetic inheritance.

Infectious agents and epigenetic mechanisms of carcinogenesis. There is growing evidence that different infectious agents may promote carcinogenesis through epigenetic mechanisms. For example, evidence has emerged that infection by HBV virus, a major risk factor for developing liver cancer, promotes hepatocarcinogenesis by inducing epigenetic changes (30,31). HBV encoded protein X (HBx) acts as an oncogenic transcription factor by affecting the expression of important cellular genes. In HBV infection-associated cancer cells HBx protein was found to upregulate expression of the DNMT genes and transcriptional silencing of key cellular genes has been attributed to promoter hyper-

For Peer Review

Downloaded from http://carcin.oxfordjournals.org/ at Imperial College London Library on June 15, 2013

methylation mediated by DNMTs (32-35). This notion is further supported by the finding that HBx also directly interacts with the de novo methyltransferase DNMT3A, directing their recruitment to specific genes and thus affecting their methylation and silencing (36). Interestingly, HBx was also shown to mediate dissociation of DNMT3A from the promoter of a set of genes, thus resulting in their hypomethylation and transcriptional activation (36). However, ChIP experiments failed to find recruitment of HBx on the activated gene-promoters suggesting that HBx could facilitate displacement of DNMT through an indirect mechanism.

In addition, a histone deacetylase (HDAC) was found to be a direct interacting partner of HBx protein, revealing a potential alternative epigenetic mechanism for its transcriptional suppressive activities (36). Furthermore, HBx was also shown to directly interact with histone acetyltransferase complex CBP/P300 and HBx-mediated recruitment of CBP/P300 complex resulting in hyperacetylation of local chromatin and transactivation of the target cellular genes (36-40). HBx was also found to downregulate the expression of E-cadherin, the gene frequently found silenced in liver cancer, by the recruitment of the mSin3A/histone deacetylase complex to the E-cadherin gene (41). Therefore, in addition to DNA methylation changes, histone deacetylation of key cellular genes may be an important epigenetic mechanism contributing to HBV-related carcinogenesis. Interestingly, in infected cells HBV genome remains in minichromosomes (small chromatin-like structures composed of histones and non-histone proteins and additional genetic material that replicate autonomously) and HBx protein was shown to regulate transcription of both viral genes through epigenetic mechanism. (42-44). However, whether the HBx-mediated engagement of cellular epigenetic machineries in replication and transcription of the viral minichromosomes on critical processes of the host cell remains to be investigated. While epigenetic mechanisms involved in carcinogenesis associated with viruses other than HBV are less understood, recent studies have suggested that human oncogenic viruses (including Epstein–Barr virus, simian virus 40, Kaposi's sarcoma-associated herpes virus, HCV virus) may be involved in the deregulation of epigenetic modifiers, ultimately resulting in deregulation of the host genes [6,7,9]. Therefore, human oncogenic viruses, in general, may promote carcinogenesis through different epigenetic mechanisms.

H. pylori is another example of an infectious agent that might promote carcinogenesis through epigenetic mechanism (45). H. pylori-infected individuals with chronic gastritis have a significantly higher risk of developing gastric cancer (46) and studies on animal models

For Peer Review

Downloaded from http://carcin.oxfordjournals.org/ at Imperial College London Library on June 15, 2013

demonstrated that infiltration of macrophages and expression of inflammation-related genes (proinflammatory cytokines) are associated with DNA methylation changes (47). Further mechanistic studies showed that DNA methylation changes associated with *H pylori*–induced chronic inflammation could be suppressed when the infected animals are treated with an anti-inflammatory agent, despite the fact that the presence of *H pylori* in gastric mucosae is unaffected (47). These findings argue that the mechanism by which some infectious agents promote carcinogenesis may involve DNA methylation changes induced by inflammation-induced mediators (45,48).

Potential modes of actions for epigenetic carcinogens

Generally, epigenetic carcinogens may promote carcinogenesis (i) through inducing direct changes in the epigenome or (ii) through an indirect deregulation of the epigenetic states. The agents that directly interact with and modulate methyl-cytosine or histone marks may be considered as direct epigenetic carcinogens. Nickel chloride, which was shown to induce changes in histone marks (49), may be considered as a direct epigenetic carcinogen.

The agents in the second group (indirect epigenetic carcinogens) include those that alter either the expression or the activity of enzymes involved in establishing and maintaining epigenetic patterns. The agents that deregulate the activity of *de novo* DNA methylation (DNMT3A and DNMT3B) or the DNA maintenance methyltransferase (DNMT1) can be considered as indirect epigenetic carcinogens. In addition, this group of agents should include the agents capable of altering the activity of proteins and protein complexes responsible for histone modifications, such as HATs and HDACs. Consistent with the evidence that HATs and histone acetylation are involved in the process of DNA repair (50), the agents that inhibit HAT and HDAC activities may compromise critical cellular process and consequently compromise genomic stability. Reduced levels of histone acetylation or enhanced histone deacetylation may result in the compaction of chromatin, blocking access of DNA repair factors to DNA lesions. Therefore, epigenetic carcinogens may transiently alter chromatin-modifying/remodelling activities, thus impeding DNA repair and other chromatin-based processes. Consistent with this notion, regional mutation rates in cancer genome was found to be largely influenced by chromatin organisation (51).

Because there is intimate and mutually reinforcing cross-talk between the three epigenetic

For Peer Review

Downloaded from http://carcin.oxfordjournals.org/ at Imperial College London Library on June 15, 2013

mechanisms in setting up and maintaining the genome-wide expression programme, epigenetic carcinogens that affect one of the interdependent epigenetic mechanisms are likely to also deregulate the other layers of the epigenome.

Our search of the Monograph Volume 100 revealed that epigenetic deregulation has been considered as a potential mechanism of carcinogenesis for only a handful of carcinogens (Table V). Yet, our comprehensive review of experimental evidence in the literature provided strong arguments for an important role of epigenetic deregulation in the mechanism by which many Group 1 carcinogens may contribute to carcinogenesis (Tables I-IV). This discrepancy may be explained by the fact that most mechanistic data on carcinogen-induced epigenome deregulation have been generated since the Volume 100 evaluation, but also by the need to resolve several issues before epigenetic testing can be fully incorporated into carcinogen evaluation. Less understanding of the relevance of epigenetic alterations and more familiarity with genetic changes as a prevailing mechanism of carcinogenesis may also account for epigenetic changes being overlooked as an important mechanism targeted by human carcinogens. Ongoing and future studies in environmental epigenetics and epigenetic toxicology may prove critical in providing crucial insights into the epigenetic mechanisms by which environmental carcinogens contribute to cancer development and progression.

Current in vivo and in vitro technologies to detect epigenetic carcinogens.

Animal models represent an essential tool in carcinogen evaluation, and they are critical for epigenetic testing. However, in addition to the complex nature of the epigenome among species and across different tissues, the lack of appropriate animal models represents the major obstacle in studying the impact of environmental carcinogens on the epigenome. Several animal models have been used to show that exposure to environmental agents has an impact on epigenetic states. These include mouse, rat, rabbit, *Drosophila*, *Caenorhabditis elegans* and zebrafish. A few mouse models harbouring locus-specific reporters have been developed and exploited in studying environmentally induced epigenetic changes. For example, the yellow *agouti* mouse model has been used as a sensitive indicator of locus-specific epigenetic changes, and several studies have demonstrated its utility in studying the role of nutritional modulation on epigenetic states (52-54). The coat colour of the *agouti* (A^{vy}) mice may be used as a sensitive read-out of

For Peer Review

Downloaded from http://carcin.oxfordjournals.org/ at Imperial College London Library on June 15, 2013

locus-specific DNA methylation states (Figure 2). This model proved to be instrumental in investigating the influence of maternal diet during pregnancy on the phenotype of the offspring (53,55). Specifically, it was found that feeding pregnant female mice a diet supplemented with folic acid, vitamin B12 or choline results in noticeable changes in the coat colour of their offspring (53). The loci responsible for the variable phenotype (coat colour) and which are susceptible to modulation by environmental exposures are known as metastable epialleles.

In addition to the *agouti* (A^{vy}) model, another murine metastable epiallele, *axin fused* (*Axin(Fu)*) (56), has been shown to exhibit epigenetic plasticity in response to changes in maternal nutrition (57). Furthermore, a model with a new metastable epiallele, CDK5 activator binding protein intracisternal A particle ($Cabp^{IAP}$), has been characterized (58). Identification of new metastable epialleles suggests that this mechanism operates at multiple loci across the mouse genome. However, the locus activity and phenotype associated with metastable mouse models are dictated by the DNA methylation pattern at the specific regulatory elements (retrotransposons) of the locus. Therefore, phenotypes associated with modulation of epialleles are locus-specific and cannot be easily extrapolated to the rest of the genome.

Despite a wide range of studies that have used these models, their unequivocal utility has been demonstrated only in the context of mother-to-offspring transmission of dietary/environmental cues. The models with metastable epialleles are also likely to contribute to characterizing the mechanism underlying transgenerational epigenetic inheritance. Current and future studies have yet to demonstrate the utility of these models in mechanistic studies aimed at identifying and evaluating carcinogens. More recently, metastable epialleles have been identified in humans (59), and with the completion of major epigenome profiling initiatives many more metastable epialleles are likely to be identified. However, metastable epialleles are likely to represent a tiny fraction of the epigenome; therefore, should carcinogens modulate these alleles, these changes may not be considered as representative of the entire epigenome.

In addition to the mouse models discussed above, rats can be considered for epigenetic testing of carcinogens. The rat has been extensively characterized for transgenerational inheritance, and there is considerable knowledge of rat embryonic and postnatal development. Therefore, the rat may be a good model for assessing the role of environmentally induced epigenetic deregulation on development and teratogenicity as

For Peer Review

Downloaded from http://carcin.oxfordjournals.org/ at Imperial College London Library on June 15, 2013

well as transgenerational epigenetic inheritance (60). However, several drawbacks associated with rat models outweigh these advantages. The genome and epigenome are far better characterized in mice than in rats, and tools are available for analyzing genome-wide changes in epigenetic states in the mouse. Therefore, the mouse represents a highly tractable model for epigenetic carcinogenicity testing. Among other *in vivo* model systems, *C. elegans*, *Drosophila*, zebrafish and honeybees offer the potential to be used in epigenetic evaluation of environmental carcinogens. For example, the fruit fly Drosophila melanogaster, a classic model for genetic research, has recently been suggested as a potential epigenetic model organism (61). In this regard, position effect variegation, the change in phenotype resulting from the change of a gene's position in the genome, was first discovered through observations of eye colour. Therefore, Drosophila eye color may serve as an attractive read-out in studying epigenetic deregulation in response to carcinogenic exposure. However, a recent study using genome-scale sequencing at single-base resolution revealed that the genome of Drosophila lacks detectable DNA methylation patterns (62), consistent with the notion that DNA methylation is dispensable for some eukaryotic organisms. Similarly, the nematode worm C. elegans completely lacks genomic DNA methylation (63). Because both C. elegans and Drosophila have been used extensively in developmental biology these organisms may represent suitable models in high-throughput screening approaches where epigenetic mechanisms (histone modifications and non-coding RNAs) can be studied independent of DNA methylation.

Finally, the use of mammalian cells (human and rodent) grown in culture should also be considered for epigenetic testing. Among these *in vitro* models, the use of stem cells (embryonic and tissue-specific) may prove particularly informative. Cell lines may prove most valuable in untargeted screening and identification of potential epigenetic carcinogens. *In vitro* models may also be instrumental in a focused dissection of epigenetic events associated with carcinogen exposure, such as molecular pathways analysis and identification of gene targets. Two major shortcomings of cellular models are their susceptibility to epigenetic alterations during *in vitro* culture and their incompatibility with transgenerational assessment, neither of which are limitations of *in vivo* models.

Genomic regions and epigenetic modifications and their biological consequences

A distinguishing feature of epigenetic changes that needs to be considered in carcinogen

For Peer Review

Downloaded from http://carcin.oxfordjournals.org/ at Imperial College London Library on June 15, 2013

evaluation is their intrinsic reversibility. Therefore, although epigenetic changes are generally stable and are usually transmitted with extreme fidelity over many cell generations, it is possible that adverse changes in DNA methylation, histone modifications and expression of non-coding RNAs associated with carcinogen exposure may be modulated by subsequent exposure to other epigenome-modulating agents.

Although intrinsic to all cells and absolutely essential, the plasticity and reversibility of the epigenome represents an important challenge in the development of methodologies and a battery of assays for epigenetic testing. In principle, epigenetic assays can be developed at two levels: (i) epigenetic patterns or marks and (ii) the expression and/or activity of epigenetic players (enzymes and other molecules involved in setting up and erasing epigenetic marks). Many environmental chemicals are likely to interfere with the activity of epigenetic players, some of which may contribute to carcinogenesis. While changes in the protein levels and activity of epigenetic machinery players (e.g. DNMT and HDAC enzymes) can be accurately measured, establishing whether the change in the expression or activity represents a lasting adverse effect with phenotypic consequences or merely an adaptive response is more challenging. Similarly, changes in epigenetic patterns (levels and patterns of DNA methylation and histone modification marks) can be detected with increasing ease and accuracy. In both cases, it is critical to establish whether a given change in either of these two epigenetic layers is anchored to a phenotypic outcome. Nevertheless, assays that monitor the epigenetic patterns should be strongly preferred as they are more likely to detect lasting changes in the epigenome.

Another important consideration is that epigenetic changes that are identified in response to a specific carcinogen should be combined with assays that establish a causative or correlative link with adverse phenotype. In this regard, it has been suggested that classifying epigenetic-based phenotypes based on the gene pathway and ontology databases may be a valuable approach (64). For example, the Gene Ontology database that is being developed with the aim of describing genes and pathways in terms of their molecular functions and associated biological processes should be a useful tool in carcinogen evaluation. An important feature of this facility is that it should allow properties to be assigned to genes and pathways in a species-independent manner, although in terms of epigenetic testing an ideal database should be exclusively focused on epigenetic effects and sets of genes and pathways that are known to be influenced by epigenetic mechanisms.

For Peer Review

Downloaded from http://carcin.oxfordjournals.org/ at Imperial College London Library on June 15, 2013

It is well established that epigenetic patterns are species- and tissue-specific; however, epigenetic changes are likely to also be exposure-specific. Because the epigenome is considered as a cellular mechanism that responds to environmental exposures, many changes in epigenetic states are likely to be a consequence of an adaptive response to adverse exposures. The key question is how to distinguish adaptive from adverse responses in the epigenome. In order to detect adverse epigenetic changes, it is essential to have a comprehensive and detailed database for comparison. With the ongoing international efforts (such as the International Human Epigenome Consortium, http://ihec-epigenomes.net), it is hoped that the epigenomes (DNA methylomes) of a wide range of human and other cell types will be available, as well as the normal variability of the epigenome. The fact that the field of epigenetics is rapidly advancing means that in the near future we are likely to be in a position to construct a "normal" reference epigenome for comparison and identification of adverse effects of potential carcinogens. Nevertheless, experiments aimed at evaluating the epigenetic effect of a carcinogen should be carefully designed to include the appropriate control group that contains a normal epigenome for comparison.

Critical technological and biological issues in assessment of epigenetic carcinogens

In recent years, we have witnessed an emergence of powerful technologies in epigenetics and epigenomics that allow the sensitive, high-throughput and genome-wide detection of epigenetic changes in normal and cancer cells (65-67). These advances, notably those linked to the development and application of microarrays and massively parallel sequencing technologies, have accelerated epigenomic research and opened up new perspectives. A wide range of methods and approaches exist for the identification, quantification and mapping of changes in the epigenome. While the earliest approaches were mostly qualitative, typically non-specific and at best useful for quantification of total epigenetic marks in cells, this field has seen considerable progress and development over the past decade.

Methods for DNA methylation analysis differ in their coverage and sensitivity, and the method of choice depends on the intended application and desired level of information. These methods include global methyl-cytosine content, degree of methylation at specific loci or genome-wide methylation maps. With the advent of more advanced and cost-

For Peer Review

Downloaded from http://carcin.oxfordjournals.org/ at Imperial College London Library on June 15, 2013

effective technologies, notably DNA microarray platforms and massively parallel sequencing, it is possible to generate comprehensive maps of epigenomes with relative ease. Similarly, a wide range of robust and genome-wide approaches has been developed for analysis of histone modifications and non-coding RNAs. These technological advances will be instrumental in establishing the epigenome in normal and diseased tissues. Considering the intimate cross-talk between the different epigenetic mechanisms, there may be value in designing approaches that aim to interrogate all layers of the epigenome (DNA methylation, histone modifications and non-coding RNA-mediated gene silencing) in response to carcinogen exposure.

Despite the remarkable progress in epigenomics, challenges still remain with regard to the analysis and interpretation of the large data sets generated by the new sequencing platforms. An important challenge will be establishing the “normal” state and the dynamic variation of the epigenome. A comprehensive understanding of the physiological variation of epigenetic states in different cell types will be critical to the capacity to discern between normal and abnormal epigenetic patterns. The development of new bioinformatics tools and epigenetic databases should facilitate these efforts (68).

Despite their genome-wide coverage, high resolution and cost-effectiveness, most methods for epigenome analysis are not compatible with the analysis of cell subpopulations. It is particularly problematic when specific epigenetic effects, similarly to genetic or other molecular effects, in mixed populations of cells need to be resolved (69). This difficulty is exemplified by the heterogeneous nature of tumour tissues, where the presence of normal (“contaminating”) cells is common.

In applying epigenetic testing, exposure to carcinogens may induce changes in specific target subpopulations of cells (e.g. stem/progenitor cells) in a tissue. Due to an inability to resolve the epigenome of a single cell and the inevitable averaging of epigenetic data, carcinogen-induced effects may be masked by tissue heterogeneity. Therefore, in addition to the capacity to map the epigenome in great detail, further effort should focus on developing robust methods capable of isolating specific cell fractions and analyzing the epigenome of single cells. Although previously established protocols can be adapted, this area is in need of further development. In addition, significant attention should be paid to selecting appropriate cell populations or subpopulations for epigenetic testing. In this regard, the use of embryonic and tissue-specific stem/progenitor cells should be considered.

For Peer Review

Downloaded from http://carcin.oxfordjournals.org/ at Imperial College London Library on June 15, 2013

Conclusions and perspectives

Unlike the genetic code, which is virtually the same in every single cell of an organism, the epigenetic code shows wide-ranging variability across different cell types and also in the same cells at different developmental stages and under the influence of various environmental stimuli. This plasticity of the epigenetic code poses a significant challenge in epigenetic testing. Current epigenetic studies, including major international sequencing projects, are expected to generate information for establishing the “normal” epigenome of tissues and cell types as well as the physiological variability of the epigenome. This should facilitate studies focusing on individual carcinogens and adverse epigenetic effects associated with carcinogen exposure. It is noteworthy that more broadly than the categories of agents considered to date in the Monographs, others such as obesity, physical inactivity and stress for example, are also likely to act through epigenetic mechanisms.

It is anticipated that chemical compounds may be classified as epigenetic carcinogens based on mechanistic evidence (in a development analogous to that applied to aristolochic acid, which was upgraded to a Group 1 carcinogen based on mechanistic data) (70). Therefore, the mechanistic epigenetic data may be used for non-genotoxic and non-receptor-mediated carcinogens (such as arsenic) or when the carcinogenicity data in humans are inconclusive. Identifying a priority set of potential epigenetic carcinogens (e.g., those classified by IARC as probably carcinogenic or possibly carcinogenic to humans [Groups 2A and 2B] that are known to be non-genotoxic) to be addressed in a systematic way may be a good starting point.

Several issues need to be resolved before epigenetic testing can be fully incorporated into carcinogen identification and evaluation and eventually used in policy decision-making. In order to incorporate epigenetic data into carcinogen evaluation, it is necessary to consider which epigenetic marks are evaluated, which assays and model systems are used and how changes in the epigenome are interpreted in terms of their potential to contribute to carcinogenesis. The cellular heterogeneity of tumour and normal tissues represents an important challenge for the interpretation of epigenomic data. Because normal or tumour tissues are rarely composed of an identical cell type or clone, tissue purity needs to be carefully considered for accurately measuring epigenetic changes associated with specific

For Peer Review

Downloaded from http://carcin.oxfordjournals.org/ at Imperial College London Library on June 15, 2013

exposures. For example, in analysing and interpreting epigenetic data associated with exposures, one may consider applying a recently developed set of analytical tools (68) for inferring changes in the distribution of different cell subpopulations using DNA methylation signatures, the methods that circumvent the need for extensive flow cytometry sorting (71,72).

Next-generation sequencing will allow remarkable accuracy, sensitivity and deep read coverage of epigenetic changes; however, extending this approach to a defined cell subpopulation or to the single-cell level remains a challenge. In addition, almost all of epigenetic profiling studies have been focused on the identification of epigenetic changes associated with annotated genes. Further efforts aimed at improving our understanding of the functional impact of aberrant epigenetic changes occurring in non-genic regions of the genome will require a comprehensive analytical methodology capable of integrating epigenomic data and transcriptomic data as well as genetic data. The development and exploitation of the Gene Ontology database and the ENCyclopedia Of DNA Elements (ENCODE; nature.com/encode) should help in this task.

Understanding the epigenetic mechanisms by which epigenetic carcinogens promote cancer development will require bringing together dedicated teams, not only of molecular biologists, toxicologists, pathologists and oncologists but also of bioinformatics and computational experts, in order to establish epigenetic assays and translate epigenomic data into an efficient and systematic evaluation of carcinogens.

Despite these significant challenges, the remarkable advances in epigenetics have provided important insights into the epigenetic mechanisms underlying carcinogenesis. The inclusion of epigenetics in the agenda of the IARC workshops on “Tumour Concordance and Mechanisms of Carcinogenesis”, held in April and November 2012, is a testimony to a growing recognition of the importance of epigenetic mechanisms in carcinogenesis. Incorporating epigenetic mechanisms into carcinogen identification and evaluation and risk assessment will be an important legacy of the IARC Monographs Programme.

For Peer Review

Downloaded from http://carcin.oxfordjournals.org/ at Imperial College London Library on June 15, 2013

Acknowledgements

We are grateful to Karen Müller for editing the manuscript. M.P. Lambert is supported by a PhD fellowship from l'Association pour le Recherche Contre le Cancer (l'ARC). The authors are supported by European Union grants No. 260791 entitled ''Eurocanplatform: A European Platform for Translational Cancer Research''; No. 308610 entitled "Exposomics: Enhanced exposure assessment and omic profiling for high priority environmental exposures in Europe"; and a French National Cancer Institute (INCa) grant "Biomarkers of B vitamins, Epigenome, genetic polymorphisms, and breast cancer risk in the European Prospective Investigation into Cancer and Nutrition (EPIC) Study".

For Peer Review

Downloaded from http://carcin.oxfordjournals.org/ at Imperial College London Library on June 15, 2013

Figure legend

Figure 1. Epigenetic mechanisms regulate key cellular processes (such as gene transcription, DNA repair, and differentiation) and play critical roles in cellular responses to environmental exposures and endogenous stimuli. Deregulation of epigenetic mechanisms may promote the development of abnormal phenotypes and cancer. There is a cross-talk between epigenetic and genetic changes in the process of cancer development and progression. Given that epigenetic and genetic changes co-exist in all cancers, it is often unclear what are the primary and secondary events pertinent to carcinogenesis.

Figure 2. Testing the effects of maternal diet on the phenotype of the offspring using the *Avy*/*a* mouse model. (A) The effect of dietary supplementation of female mice before or after mating with male Avy/a agouti mice can be tested. (B) The effect of maternal dietary supplementation on the epigenome can be "read" by coat-colour distribution in Avy/a offspring. (C) The effect of DNA methylation at the intracisternal A particle (IAP) on agouti gene expression. Maternal dietary supplementation may shift the coat-colour distribution of the offspring. This shift is result of change in methylation at IAP element upstream of the agouti gene. Adapted from (73).

For Peer Review

Downloaded from http://carcin.oxfordjournals.org/ at Imperial College London Library on June 15, 2013

Table I. Epigenetic mechanisms and Volume 100A carcinogens

Human carcinogen	Epigenetic mechanism	Model			References
		Human samples	Cell lines	Animal model	
Part A: Pharmaceuticals					
Tamoxifen					
	DNA methylation	*			(74,75)
			*		(76)
				*	(77,78)
	Histone marks			*	(78)
	miRNA			*	(79)
Diethylstilbestrol					
	DNA methylation		*		(80)
				*	(80-86)
	Histone marks		*		(87)
	miRNA	*			(88)
			*		(88)
Prostaglandin E2					
	DNA methylation		*		(89,90)
				*	(89)
	Histone marks		*		(91)
	miRNA		*		(92,93)
Hormone Therapy					
	DNA methylation	*			(94-96)
Cyclophosphamide					
	miRNA			*	(97)
Chromium VI compounds					
	Histone marks		*		(98,99)

For Peer Review

Downloaded from http://carcin.oxfordjournals.org/ at Imperial College London Library on June 15, 2013

Table II. Epigenetic mechanisms and Volume 100B carcinogens

Human carcinogen	Epigenetic mechanism	Model			References
		Human samples	Cell lines	Animal model	
Part B: Biological agents					
Hepatitis B virus					
	DNA methylation	*			(38,100-110)
			*		(36,38,101,110-114)
	Histone marks	*			(115,116)
			*		(39,41,115-117)
				*	(115,116)
	miRNA	*			(118-123)
			*		(41,118-120,124)
				*	(124)
Hepatitis C virus					
	DNA methylation	*			(105,106,125-127)
			*		(128,129)
				*	(130)
	Histone marks		*		(131,132)
	miRNA	*			(123,133-136)
			*		(135,137-141)
Human Papillomaviruses					
	DNA methylation	*			(142,143)
			*		(142,144)
	Histone marks		*		(145-150)
	miRNA		*		(151-153)
Human T-cell Lymphotropic Virus Type 1					
	DNA methylation		*		(154)
	Histone marks		*		(155-157)
	miRNA	*			(158,159)
			*		(159,160)
Epstein–Barr virus					
	DNA methylation	*			(161-170)
			*		(161-164,171-173)
	Histone marks		*		(171,174-178)
	miRNA	*			(179-181)
			*		(122,179,181-192)
				*	(181)
Helicobacter pylori					
	DNA methylation	*			(115,116,193-205)

For Peer Review

Downloaded from http://carcin.oxfordjournals.org/ at Imperial College London Library on June 15, 2013

			*		(193-195,198,206-208)
				*	(205,207)
	Histone marks		*		(206,209-212)
	miRNA	*			(213)
			*		(213)

For Peer Review

Downloaded from http://carcin.oxfordjournals.org/ at Imperial College London Library on June 15, 2013

Table III. Epigenetic mechanisms and Volume 100C-D carcinogens

Human carcinogen	Epigenetic mechanism	Model			References
		Human samples	Cell lines	Animal model	
Part C : Arsenic, Metals, Fibres, and Dusts and Part D: Radiation					
Arsenic					
	DNA methylation	*			[176-182]
			*		[183-188]
				*	[183,189,190]
	Histone marks		*		[191]
	miRNA		*		[192]
Cadmium					
	DNA methylation		*		[193-196]
				*	[197-200]
	miRNA	*			[201]
Nickel					
	DNA methylation		*		[202,203]
				*	[74]
	Histone marks		*		[20,204-214]
				*	[214]
Beryllium					
	DNA methylation			*	(214)
Asbestos					
	DNA methylation			*	(215)
	miRNA			*	(216)
X-radiation					
	DNA methylation	*			(217)
				*	(218-220)
	miRNA	*			(221)
			*		(222)
				*	(220)
Gamma radiation					
	DNA methylation		*		(223)
	miRNA			*	(221,224,225)
Smoky coal emissions (for cooking and heating)					
	DNA methylation	*			(226)
Chromium VI compounds					
	Histone marks		*		[233,234]

Review

Downloaded from http://carcin.oxfordjournals.org/ at Imperial College London Library on June 15, 2013

Table IV. Epigenetic mechanisms and Volume 100E carcinogens

Human carcinogen	Epigenetic mechanism	Model			References
		Human samples	Cell lines	Animal model	
Lifestyle and diet					
Tobacco smoking					
	DNA methylation	*			(110,227-236)
			*		(227,228,237-239)
	Histone marks		*		(240)
Alcohol consumption					
	DNA methylation	*			(241-244)
			*		(245,246)
				*	(247-249)
	Histone marks		*		(245,250)
				*	(193,194,251)
	miRNA	*			(81,252,253)

For Peer Review

Downloaded from http://carcin.oxfordjournals.org/ at Imperial College London Library on June 15, 2013

Table V. Epigenetic mechanisms and Volume 100 carcinogens

Agent	Major mechanism	2nd mechanism	Human cancer	Comment
Diethystilbestrol	Estrogen receptor-dependent pathway	Epigenetic reprogramming	Cervix, vagina, breast, endometrium (limited evidence), testis (limited evidence)	
Hepatitis B virus	Integration into host DNA		Hepatocellular carcinoma, cholangiocarcinoma (limited evidence), non-Hodgkin lymphoma (limited evidence)	Epigenetc silencing of tumour suppressor genes; interaction with aflatoxins
Chromium VI compounds	Genotoxicity	Epigenetic effects	Lung, nasal cavity and paranasal sinuses (limited evidence)	
Ionizing radiation	Energy transfer in clusters	Epigenetic effects	Many	
Coke production	Genotoxicity	Epigenetic effects	Lung	

For Peer Review

Downloaded from http://carcin.oxfordjournals.org/ at Imperial College London Library on June 15, 2013

References

1. Bell, J.T., *et al.* (2011) A twin approach to unraveling epigenetics. *Trends in genetics : TIG*, **27**, 116-25.
2. Christensen, B.C., *et al.* (2009) Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. *PLoS Genet*, **5**, e1000602.
3. Hou, L., *et al.* (2012) Environmental chemical exposures and human epigenetics. *International journal of epidemiology*, **41**, 79-105.
4. Feil, R., *et al.* (2011) Epigenetics and the environment: emerging patterns and implications. *Nature reviews. Genetics*, **13**, 97-109.
5. Baccarelli, A., *et al.* (2009) Epigenetics and environmental chemicals. *Current opinion in pediatrics*, **21**, 243-51.
6. Herceg, Z. (2007) Epigenetics and cancer: towards an evaluation of the impact of environmental and dietary factors. *Mutagenesis*, **22**, 91-103.
7. Herceg, Z., *et al.* (2011) Epigenetic mechanisms and cancer: an interface between the environment and the genome. *Epigenetics : official journal of the DNA Methylation Society*, **6**, 804-19.
8. Jaenisch, R., *et al.* (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet*, **33**, 245-54.
9. Milosavljevic, A. (2011) Emerging patterns of epigenomic variation. *Trends in genetics : TIG*, **27**, 242-50.
10. Biron, V.L., *et al.* (2004) Distinct dynamics and distribution of histone methyl-lysine derivatives in mouse development. *Developmental biology*, **276**, 337-51.
11. Barouki, R., *et al.* (2012) Developmental origins of non-communicable disease: implications for research and public health. *Environmental health : a global access science source*, **11**, 42.
12. Reik, W. (2007) Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature*, **447**, 425-32.
13. Mayer, W., *et al.* (2000) Demethylation of the zygotic paternal genome. *Nature*, **403**, 501-2.
14. Gluckman, P.D., *et al.* (2008) Effect of in utero and early-life conditions on adult health and disease. *N Engl J Med*, **359**, 61-73.
15. Skinner, M.K. (2011) Environmental epigenetic transgenerational inheritance and somatic epigenetic mitotic stability. *Epigenetics : official journal of the DNA Methylation Society*, **6**, 838-42.
16. Anway, M.D., *et al.* (2005) Epigenetic transgenerational actions of endocrine disruptors and male fertility. *Science*, **308**, 1466-9.
17. Smith, Z.D., *et al.* (2013) DNA methylation: roles in mammalian development. *Nature reviews. Genetics*, **14**, 204-20.
18. Klaunig, J.E., *et al.* (2000) Epigenetic mechanisms of chemical carcinogenesis. *Human & experimental toxicology*, **19**, 543-55.
19. Melnick, R.L., *et al.* (1996) Implications for risk assessment of suggested nongenotoxic mechanisms of chemical carcinogenesis. *Environmental health perspectives*, **104 Suppl 1**, 123-34.
20. Cogliano, V.J., *et al.* (2011) Preventable exposures associated with human cancers. *Journal of the National Cancer Institute*, **103**, 1827-39.
21. IARC-Monograph (2012) IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. **100**.
22. Giusti, R.M., *et al.* (1995) Diethylstilbestrol revisited: a review of the long-term health effects. *Annals of internal medicine*, **122**, 778-88.

For Peer Review

Downloaded from http://carcin.oxfordjournals.org/ at Imperial College London Library on June 15, 2013

23. IARC-Monograph (1987) IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. **Supplement 7**, 273.
24. Newbold, R.R., *et al.* (2006) Adverse effects of the model environmental estrogen diethylstilbestrol are transmitted to subsequent generations. *Endocrinology*, **147**, S11-7.
25. Newbold, R.R., *et al.* (2002) Characterization of uterine leiomyomas in CD-1 mice following developmental exposure to diethylstilbestrol (DES). *Toxicologic pathology*, **30**, 611-6.
26. Newbold, R.R., *et al.* (2000) Proliferative lesions and reproductive tract tumors in male descendants of mice exposed developmentally to diethylstilbestrol. *Carcinogenesis*, **21**, 1355-63.
27. Hsu, P.Y., *et al.* (2009) Xenoestrogen-induced epigenetic repression of microRNA-9-3 in breast epithelial cells. *Cancer research*, **69**, 5936-45.
28. Lee, Y.M., *et al.* (2011) miRNA-34b as a tumor suppressor in estrogen-dependent growth of breast cancer cells. *Breast cancer research : BCR*, **13**, R116.
29. Warita, K., *et al.* (2010) Direct effects of diethylstilbestrol on the gene expression of the cholesterol side-chain cleavage enzyme (P450scc) in testicular Leydig cells. *Life sciences*, **87**, 281-5.
30. Herceg, Z., *et al.* (2011) Epigenetic mechanisms in hepatocellular carcinoma: how environmental factors influence the epigenome. *Mutation research*, **727**, 55-61.
31. Arzumanyan, A., *et al.* (2013) Pathogenic mechanisms in HBV- and HCV-associated hepatocellular carcinoma. *Nature reviews. Cancer*, **13**, 123-35.
32. Calvisi, D.F., *et al.* (2007) Mechanistic and prognostic significance of aberrant methylation in the molecular pathogenesis of human hepatocellular carcinoma. *J Clin Invest*, **117**, 2713-22.
33. Huang, J., *et al.* (2010) Down-regulated microRNA-152 induces aberrant DNA methylation in hepatitis B virus-related hepatocellular carcinoma by targeting DNA methyltransferase 1. *Hepatology*, **52**, 60-70.
34. Park, S.H., *et al.* (2011) Hepatitis B virus X protein overcomes all-trans retinoic acid-induced cellular senescence by downregulating levels of p16 and p21 via DNA methylation. *The Journal of general virology*, **92**, 1309-17.
35. Um, T.H., *et al.* (2011) Aberrant CpG island hypermethylation in dysplastic nodules and early HCC of hepatitis B virus-related human multistep hepatocarcinogenesis. *Journal of hepatology*, **54**, 939-47.
36. Zheng, D.L., *et al.* (2009) Epigenetic modification induced by hepatitis B virus X protein via interaction with de novo DNA methyltransferase DNMT3A. *J Hepatol*, **50**, 377-87.
37. Horikawa, I., *et al.* (2003) Transcriptional regulation of the telomerase hTERT gene as a target for cellular and viral oncogenic mechanisms. *Carcinogenesis*, **24**, 1167-76.
38. Park, I.Y., *et al.* (2007) Aberrant epigenetic modifications in hepatocarcinogenesis induced by hepatitis B virus X protein. *Gastroenterology*, **132**, 1476-94.
39. Shon, J.K., *et al.* (2009) Hepatitis B virus-X protein recruits histone deacetylase 1 to repress insulin-like growth factor binding protein 3 transcription. *Virus Res*, **139**, 14-21.
40. Cougot, D., *et al.* (2007) The hepatitis B virus X protein functionally interacts with CREB-binding protein/p300 in the regulation of CREB-mediated transcription. *J Biol Chem*, **282**, 4277-87.
41. Arzumanyan, A., *et al.* (2012) Epigenetic repression of E-cadherin expression by hepatitis B virus x antigen in liver cancer. *Oncogene*, **31**, 563-72.
42. Fernandez, A.F., *et al.* (2010) Viral epigenomes in human tumorigenesis. *Oncogene*, **29**, 1405-20.
43. Herceg, Z., *et al.* (2009) HBV protein as a double-barrel shot-gun targets epigenetic landscape in liver cancer. *J Hepatol*, **50**, 252-5.
44. Pollicino, T., *et al.* (2007) Molecular and functional analysis of occult hepatitis B virus isolates from patients with hepatocellular carcinoma. *Hepatology*, **45**, 277-85.

For Peer Review

Downloaded from http://carcin.oxfordjournals.org/ at Imperial College London Library on June 15, 2013

45. Chiba, T., *et al.* (2012) Inflammation-associated cancer development in digestive organs: mechanisms and roles for genetic and epigenetic modulation. *Gastroenterology*, **143**, 550-63.
46. Uemura, N., *et al.* (2001) Helicobacter pylori infection and the development of gastric cancer. *The New England journal of medicine*, **345**, 784-9.
47. Katsurano, M., *et al.* (2012) Early-stage formation of an epigenetic field defect in a mouse colitis model, and non-essential roles of T- and B-cells in DNA methylation induction. *Oncogene*, **31**, 342-51.
48. Martin, M., *et al.* (2012) From hepatitis to hepatocellular carcinoma: a proposed model for cross-talk between inflammation and epigenetic mechanisms. *Genome medicine*, **4**, 8.
49. Ke, Q., *et al.* (2006) Alterations of histone modifications and transgene silencing by nickel chloride. *Carcinogenesis*, **27**, 1481-8.
50. Gospodinov, A., *et al.* (2012) Shaping chromatin for repair. *Mutation research*, **12**.
51. Schuster-Bockler, B., *et al.* (2012) Chromatin organization is a major influence on regional mutation rates in human cancer cells. *Nature*, **488**, 504-7.
52. Duhl, D.M., *et al.* (1994) Neomorphic agouti mutations in obese yellow mice. *Nat Genet*, **8**, 59-65.
53. Wolff, G.L., *et al.* (1998) Maternal epigenetics and methyl supplements affect agouti gene expression in Avy/a mice. *Faseb J*, **12**, 949-57.
54. Waterland, R.A., *et al.* (2003) Transposable elements: targets for early nutritional effects on epigenetic gene regulation. *Mol Cell Biol*, **23**, 5293-300.
55. Cooney, C.A., *et al.* (2002) Maternal methyl supplements in mice affect epigenetic variation and DNA methylation of offspring. *J Nutr*, **132**, 2393S-2400S.
56. Rakyan, V.K., *et al.* (2003) Transgenerational inheritance of epigenetic states at the murine Axin(Fu) allele occurs after maternal and paternal transmission. *Proc Natl Acad Sci U S A*, **100**, 2538-43.
57. Waterland, R.A. (2006) Assessing the effects of high methionine intake on DNA methylation. *J Nutr*, **136**, 1706S-1710S.
58. Druker, R., *et al.* (2004) Complex patterns of transcription at the insertion site of a retrotransposon in the mouse. *Nucleic Acids Res*, **32**, 5800-8.
59. Waterland, R.A., *et al.* (2010) Season of conception in rural gambia affects DNA methylation at putative human metastable epialleles. *PLoS genetics*, **6**, e1001252.
60. Skinner, M.K., *et al.* (2012) Epigenetic transgenerational inheritance of somatic transcriptomes and epigenetic control regions. *Genome biology*, **13**, R91.
61. Lyko, F., *et al.* (2006) Epigenetic regulation in Drosophila. *Current topics in microbiology and immunology*, **310**, 23-44.
62. Raddatz, G., *et al.* (2013) Dnmt2-dependent methylomes lack defined DNA methylation patterns. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 8627-31.
63. Bird, A. (2002) DNA methylation patterns and epigenetic memory. *Genes Dev*, **16**, 6-21.
64. Goodman, J.I., *et al.* (2010) What do we need to know prior to thinking about incorporating an epigenetic evaluation into safety assessments? *Toxicological sciences : an official journal of the Society of Toxicology*, **116**, 375-81.
65. Laird, P.W. (2010) Principles and challenges of genome-wide DNA methylation analysis. *Nat Rev Genet*, **11**, 191-203.
66. Esteller, M. (2007) Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat Rev Genet*, **8**, 286-98.
67. Umer, M., *et al.* (2013) Deciphering the Epigenetic Code: An Overview of DNA Methylation Analysis Methods. *Antioxidants & redox signaling*, **12**.
68. Bock, C. (2012) Analysing and interpreting DNA methylation data. *Nature reviews. Genetics*, **13**, 705-19.

For Peer Review

Downloaded from http://carcin.oxfordjournals.org/ at Imperial College London Library on June 15, 2013

69. Herceg, Z., *et al.* (2011) New concepts of old epigenetic phenomena and their implications for selecting specific cell populations for epigenomic research. *Epigenomics*, **12**, 383-6.
70. Olivier, M., *et al.* (2012) Upper urinary tract urothelial cancers: where it is A:T. *Nature reviews. Cancer*, **12**, 503-4.
71. Houseman, E.A., *et al.* (2012) DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC bioinformatics*, **13**, 86.
72. Liu, Y., *et al.* (2013) Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nature biotechnology*, **12**.
73. Jirtle, R.L., *et al.* (2007) Environmental epigenomics and disease susceptibility. *Nat Rev Genet*, **8**, 253-62.
74. Liggett, T.E., *et al.* (2011) Methylation patterns in cell-free plasma DNA reflect removal of the primary tumor and drug treatment of breast cancer patients. *Int J Cancer*, **128**, 492-9.
75. Wu, H., *et al.* (2005) Hypomethylation-linked activation of PAX2 mediates tamoxifen-stimulated endometrial carcinogenesis. *Nature*, **438**, 981-7.
76. Pietkiewicz, P.P., *et al.* (2010) Tamoxifen epigenetically modulates CXCL12 expression in MCF-7 breast cancer cells. *Biomed Pharmacother*, **64**, 54-7.
77. Pathak, S., *et al.* (2009) Effect of tamoxifen treatment on global and insulin-like growth factor 2-H19 locus-specific DNA methylation in rat spermatozoa and its association with embryo loss. *Fertil Steril*, **91**, 2253-63.
78. Tryndyak, V.P., *et al.* (2006) Effect of long-term tamoxifen exposure on genotoxic and epigenetic changes in rat liver: implications for tamoxifen-induced hepatocarcinogenesis. *Carcinogenesis*, **27**, 1713-20.
79. Pogribny, I.P., *et al.* (2007) Induction of microRNAome deregulation in rat liver by long-term tamoxifen exposure. *Mutat Res*, **619**, 30-7.
80. Bromer, J.G., *et al.* (2009) Hypermethylation of homeobox A10 by in utero diethylstilbestrol exposure: an epigenetic mechanism for altered developmental programming. *Endocrinology*, **150**, 3376-82.
81. Tang, W.Y., *et al.* (2008) Persistent hypomethylation in the promoter of nucleosomal binding protein 1 (Nsbp1) correlates with overexpression of Nsbp1 in mouse uteri neonatally exposed to diethylstilbestrol or genistein. *Endocrinology*, **149**, 5922-31.
82. Li, S., *et al.* (2003) Neonatal diethylstilbestrol exposure induces persistent elevation of c-fos expression and hypomethylation in its exon-4 in mouse uterus. *Mol Carcinog*, **38**, 78-84.
83. Sato, K., *et al.* (2006) Neonatal exposure to diethylstilbestrol alters the expression of DNA methyltransferases and methylation of genomic DNA in the epididymis of mice. *Endocr J*, **53**, 331-7.
84. Sato, K., *et al.* (2009) Neonatal exposure to diethylstilbestrol alters expression of DNA methyltransferases and methylation of genomic DNA in the mouse uterus. *Endocr J*, **56**, 131-9.
85. Li, S., *et al.* (2001) Promoter CpG methylation of Hox-a10 and Hox-a11 in mouse uterus not altered upon neonatal diethylstilbestrol exposure. *Mol Carcinog*, **32**, 213-9.
86. Alworth, L.C., *et al.* (2002) Uterine responsiveness to estradiol and DNA methylation are altered by fetal exposure to diethylstilbestrol and methoxychlor in CD-1 mice: effects of low versus high doses. *Toxicol Appl Pharmacol*, **183**, 10-22.
87. Warita, K., *et al.* (2010) Direct effects of diethylstilbestrol on the gene expression of the cholesterol side-chain cleavage enzyme (P450scc) in testicular Leydig cells. *Life Sci*, **87**, 281-5.
88. Hsu, P.Y., *et al.* (2009) Xenoestrogen-induced epigenetic repression of microRNA-9-3 in breast epithelial cells. *Cancer Res*, **69**, 5936-45.
89. Xia, D., *et al.* (2012) Prostaglandin E2 promotes intestinal tumor growth via DNA methylation. *Nat Med*, **18**, 224-6.

For Peer Review

Downloaded from http://carcin.oxfordjournals.org/ at Imperial College London Library on June 15, 2013

90. Huang, S.K., *et al.* (2012) Prostaglandin E(2) increases fibroblast gene-specific and global DNA methylation via increased DNA methyltransferase expression. *FASEB J*, **26**, 3703-14.
91. Kawagoe, J., *et al.* (2007) Mechanism of the divergent effects of estrogen on the cell proliferation of human umbilical endothelial versus aortic smooth muscle cells. *Endocrinology*, **148**, 6092-9.
92. Wickramasinghe, N.S., *et al.* (2009) Estradiol downregulates miR-21 expression and increases miR-21 target gene expression in MCF-7 breast cancer cells. *Nucleic Acids Res*, **37**, 2584-95.
93. Maillot, G., *et al.* (2009) Widespread estrogen-dependent repression of micrornas involved in breast tumor cell growth. *Cancer Res*, **69**, 8332-40.
94. Wu, A.H., *et al.* (2010) Hormone therapy, DNA methylation and colon cancer. *Carcinogenesis*, **31**, 1060-7.
95. Friso, S., *et al.* (2007) Oestrogen replacement therapy reduces total plasma homocysteine and enhances genomic DNA methylation in postmenopausal women. *Br J Nutr*, **97**, 617-21.
96. Cappelletti, V., *et al.* (2008) Patterns and changes in gene expression following neo-adjuvant anti-estrogen treatment in estrogen receptor-positive breast cancer. *Endocr Relat Cancer*, **15**, 439-49.
97. Gueta, K., *et al.* (2010) Teratogen-induced alterations in microRNA-34, microRNA-125b and microRNA-155 expression: correlation with embryonic p53 genotype and limb phenotype. *BMC Dev Biol*, **10**, 20.
98. Xia, B., *et al.* (2011) Chromium(VI) causes down regulation of biotinidase in human bronchial epithelial cells by modifications of histone acetylation. *Toxicol Lett*, **205**, 140-5.
99. Sun, H., *et al.* (2009) Modulation of histone methylation and MLH1 gene silencing by hexavalent chromium. *Toxicol Appl Pharmacol*, **237**, 258-66.
100. Lambert, M.P., *et al.* (2011) Aberrant DNA methylation distinguishes hepatocellular carcinoma associated with HBV and HCV infection and alcohol intake. *J Hepatol*, **54**, 705-15.
101. Vivekanandan, P., *et al.* (2010) Hepatitis B virus replication induces methylation of both host and viral DNA. *J Virol*, **84**, 4321-9.
102. Oh, B.K., *et al.* (2007) DNA methyltransferase expression and DNA methylation in human hepatocellular carcinoma and their clinicopathological correlation. *Int J Mol Med*, **20**, 65-73.
103. Zhu, Y.Z., *et al.* (2010) Hepatitis B virus X protein induces hypermethylation of p16(INK4A) promoter via DNA methyltransferases in the early stage of HBV-associated hepatocarcinogenesis. *J Viral Hepat*, **17**, 98-107.
104. Zhu, Y.Z., *et al.* (2010) Hepatitis B virus X protein promotes hypermethylation of p16(INK4A) promoter through upregulation of DNA methyltransferases in hepatocarcinogenesis. *Exp Mol Pathol*, **89**, 268-75.
105. Hernandez-Vargas, H., *et al.* (2010) Hepatocellular carcinoma displays distinct DNA methylation signatures with potential as clinical predictors. *PLoS One*, **5**, e9749.
106. Feng, Q., *et al.* (2010) DNA methylation changes in normal liver tissues and hepatocellular carcinoma with different viral infection. *Exp Mol Pathol*, **88**, 287-92.
107. Su, P.F., *et al.* (2007) Differential DNA methylation associated with hepatitis B virus infection in hepatocellular carcinoma. *Int J Cancer*, **121**, 1257-64.
108. He, Y., *et al.* (2006) Not polymorphism but methylation of class II transactivator gene promoter IV associated with persistent HBV infection. *J Clin Virol*, **37**, 282-6.
109. Jicai, Z., *et al.* (2006) Persistent infection of hepatitis B virus is involved in high rate of p16 methylation in hepatocellular carcinoma. *Mol Carcinog*, **45**, 530-6.
110. Liu, J., *et al.* (2006) Downregulation of E-cadherin by hepatitis B virus X antigen in hepatocellullar carcinoma. *Oncogene*, **25**, 1008-17.

For Peer Review

Downloaded from http://carcin.oxfordjournals.org/ at Imperial College London Library on June 15, 2013

111. Niu, D., *et al.* (2009) HBx genotype D represses GSTP1 expression and increases the oxidative level and apoptosis in HepG2 cells. *Mol Oncol*, **3**, 67-76.
112. Jung, J.K., *et al.* (2007) Expression of DNA methyltransferase 1 is activated by hepatitis B virus X protein via a regulatory circuit involving the p16INK4a-cyclin D1-CDK 4/6-pRb-E2F1 pathway. *Cancer Res*, **67**, 5771-8.
113. Jung, J.K., *et al.* (2010) Hepatitis B virus X protein overcomes the growth-inhibitory potential of retinoic acid by downregulating retinoic acid receptor-beta2 expression via DNA methylation. *J Gen Virol*, **91**, 493-500.
114. Lee, J.O., *et al.* (2005) Hepatitis B virus X protein represses E-cadherin expression via activation of DNA methyltransferase 1. *Oncogene*, **24**, 6617-25.
115. Yoo, E.J., *et al.* (2008) Helicobacter pylori-infection-associated CpG island hypermethylation in the stomach and its possible association with polycomb repressive marks. *Virchows Arch*, **452**, 515-24.
116. Yoo, Y.G., *et al.* (2008) Hepatitis B virus X protein induces the expression of MTA1 and HDAC1, which enhances hypoxia signaling in hepatocellular carcinoma cells. *Oncogene*, **27**, 3405-13.
117. Yang, L., *et al.* (2009) Hepatitis B virus X protein upregulates expression of SMYD3 and C-MYC in HepG2 cells. *Med Oncol*, **26**, 445-51.
118. Zhang, J., *et al.* (2011) Methylation of RAR-beta2, RASSF1A, and CDKN2A genes induced by nickel subsulfide and nickel-carcinogenesis in rats. *Biomed Environ Sci*, **24**, 163-71.
119. Zhang, X., *et al.* (2011) Modulation of hepatitis B virus replication and hepatocyte differentiation by MicroRNA-1. *Hepatology*, **53**, 1476-85.
120. Zhang, Z.Z., *et al.* (2011) Hepatitis B virus and hepatocellular carcinoma at the miRNA level. *World J Gastroenterol*, **17**, 3353-8.
121. Li, G., *et al.* (2010) MicroRNA-10b induced by Epstein-Barr virus-encoded latent membrane protein-1 promotes the metastasis of human nasopharyngeal carcinoma cells. *Cancer Lett*, **299**, 29-36.
122. Li, L.M., *et al.* (2010) Serum microRNA profiles serve as novel biomarkers for HBV infection and diagnosis of HBV-positive hepatocarcinoma. *Cancer Res*, **70**, 9798-807.
123. Ura, S., *et al.* (2009) Differential microRNA expression between hepatitis B and hepatitis C leading disease progression to hepatocellular carcinoma. *Hepatology*, **49**, 1098-112.
124. Kong, G., *et al.* (2011) Upregulated microRNA-29a by hepatitis B virus X protein enhances hepatoma cell migration by targeting PTEN in cell culture model. *PLoS One*, **6**, e19518.
125. Deng, Y.B., *et al.* (2010) Identification of genes preferentially methylated in hepatitis C virus-related hepatocellular carcinoma. *Cancer Sci*, **101**, 1501-10.
126. Lu, C.Y., *et al.* (2009) Aberrant DNA methylation profile and frequent methylation of KLK10 and OXGR1 genes in hepatocellular carcinoma. *Genes Chromosomes Cancer*, **48**, 1057-68.
127. Nishida, N., *et al.* (2008) Aberrant methylation of multiple tumor suppressor genes in aging liver, chronic hepatitis, and hepatocellular carcinoma. *Hepatology*, **47**, 908-18.
128. Lim, J.S., *et al.* (2012) Hepatitis C virus Core protein overcomes stress-induced premature senescence by down-regulating p16 expression via DNA methylation. *Cancer Lett*, **321**, 154-61.
129. Ripoli, M., *et al.* (2011) Hypermethylated levels of E-cadherin promoter in Huh-7 cells expressing the HCV core protein. *Virus Res*, **160**, 74-81.
130. Higgs, M.R., *et al.* (2010) Downregulation of Gadd45beta expression by hepatitis C virus leads to defective cell cycle arrest. *Cancer Res*, **70**, 4901-11.
131. Miura, K., *et al.* (2008) Hepatitis C virus-induced oxidative stress suppresses hepcidin expression through increased histone deacetylase activity. *Hepatology*, **48**, 1420-9.

For Peer Review

Downloaded from http://carcin.oxfordjournals.org/ at Imperial College London Library on June 15, 2013

132. Borowski, P., *et al.* (1999) Identification and characterization of a histone binding site of the non-structural protein 3 of hepatitis C virus. *J Clin Virol*, **13**, 61-9.
133. Peveling-Oberhag, J., *et al.* (2012) Dysregulation of global microRNA expression in splenic marginal zone lymphoma and influence of chronic hepatitis C virus infection. *Leukemia*, **26**, 1654-62.
134. Abdalla, M.A., *et al.* (2012) Promising Candidate Urinary MicroRNA Biomarkers for the Early Detection of Hepatocellular Carcinoma among High-Risk Hepatitis C Virus Egyptian Patients. *J Cancer*, **12**, 19-31.
135. Marquez, R.T., *et al.* (2010) Correlation between microRNA expression levels and clinical parameters associated with chronic hepatitis C viral infection in humans. *Lab Invest*, **90**, 1727-36.
136. Peng, X., *et al.* (2009) Computational identification of hepatitis C virus associated microRNA-mRNA regulatory modules in human livers. *BMC Genomics*, **10**, 373.
137. Cermelli, S., *et al.* (2011) Circulating microRNAs in patients with chronic hepatitis C and non-alcoholic fatty liver disease. *PLoS One*, **6**, e23937.
138. Ishida, H., *et al.* (2011) Alterations in microRNA expression profile in HCV-infected hepatoma cells: involvement of miR-491 in regulation of HCV replication via the PI3 kinase/Akt pathway. *Biochem Biophys Res Commun*, **412**, 92-7.
139. Steuerwald, N.M., *et al.* (2010) Parallel microRNA and mRNA expression profiling of (genotype 1b) human hepatoma cells expressing hepatitis C virus. *Liver Int*, **30**, 1490-504.
140. Braconi, C., *et al.* (2010) Hepatitis C virus proteins modulate microRNA expression and chemosensitivity in malignant hepatocytes. *Clin Cancer Res*, **16**, 957-66.
141. Liu, X., *et al.* (2010) Systematic identification of microRNA and messenger RNA profiles in hepatitis C virus-infected human hepatoma cells. *Virology*, **398**, 57-67.
142. Sartor, M.A., *et al.* (2011) Genome-wide methylation and expression differences in HPV(+) and HPV(-) squamous cell carcinoma cell lines are consistent with divergent mechanisms of carcinogenesis. *Epigenetics*, **6**, 777-87.
143. Weiss, D., *et al.* (2011) Promoter methylation of cyclin A1 is associated with human papillomavirus 16 induced head and neck squamous cell carcinoma independently of p53 mutation. *Mol Carcinog*, **50**, 680-8.
144. Jiang, J., *et al.* (2012) Hypomethylated CpG around the transcription start site enables TERT expression and HPV16 E6 regulates TERT methylation in cervical cancer cells. *Gynecol Oncol*, **124**, 534-41.
145. Hyland, P.L., *et al.* (2011) Evidence for alteration of EZH2, BMI1, and KDM6A and epigenetic reprogramming in human papillomavirus type 16 E6/E7-expressing keratinocytes. *J Virol*, **85**, 10999-1006.
146. McLaughlin-Drubin, M.E., *et al.* (2011) Human papillomavirus E7 oncoprotein induces KDM6A and KDM6B histone demethylase expression and causes epigenetic reprogramming. *Proc Natl Acad Sci U S A*, **108**, 2130-5.
147. Bodily, J.M., *et al.* (2011) Human papillomavirus E7 enhances hypoxia-inducible factor 1-mediated transcription by inhibiting binding of histone deacetylases. *Cancer Res*, **71**, 1187-95.
148. Longworth, M.S., *et al.* (2005) HPV31 E7 facilitates replication by activating E2F2 transcription through its interaction with HDACs. *EMBO J*, **24**, 1821-30.
149. Avvakumov, N., *et al.* (2003) Interaction of the HPV E7 proteins with the pCAF acetyltransferase. *Oncogene*, **22**, 3833-41.
150. Hsu, C.H., *et al.* (2012) The HPV E6 oncoprotein targets histone methyltransferases for modulating specific gene transcription. *Oncogene*, **31**, 2335-49.
151. Dreher, A., *et al.* (2011) Differential expression of cellular microRNAs in HPV 11, -16, and -45 transfected cells. *Biochem Biophys Res Commun*, **412**, 20-5.

For Peer Review

Downloaded from http://carcin.oxfordjournals.org/ at Imperial College London Library on June 15, 2013

152. Greco, D., *et al.* (2011) Human papillomavirus 16 E5 modulates the expression of host microRNAs. *PLoS One*, **6**, e21646.
153. Martinez, I., *et al.* (2008) Human papillomavirus type 16 reduces the expression of microRNA-218 in cervical carcinoma cells. *Oncogene*, **27**, 2575-82.
154. Nakase, K., *et al.* (2009) Mechanisms of SHP-1 P2 promoter regulation in hematopoietic cells and its silencing in HTLV-1-transformed T cells. *J Leukoc Biol*, **85**, 165-74.
155. Bogenberger, J.M., *et al.* (2008) Human T Lymphotropic Virus Type 1 protein Tax reduces histone levels. *Retrovirology*, **5**, 9.
156. Cheng, J., *et al.* (2007) Negative regulation of the SH2-homology containing protein-tyrosine phosphatase-1 (SHP-1) P2 promoter by the HTLV-1 Tax oncoprotein. *Blood*, **110**, 2110-20.
157. Lu, H., *et al.* (2004) Tax relieves transcriptional repression by promoting histone deacetylase 1 release from the human T-cell leukemia virus type 1 long terminal repeat. *J Virol*, **78**, 6735-43.
158. Bellon, M., *et al.* (2009) Deregulation of microRNA involved in hematopoiesis and the immune response in HTLV-I adult T-cell leukemia. *Blood*, **113**, 4914-7.
159. Yeung, M.L., *et al.* (2008) Roles for microRNAs, miR-93 and miR-130b, and tumor protein 53-induced nuclear protein 1 tumor suppressor in cell growth dysregulation by human T-cell lymphotrophic virus 1. *Cancer Res*, **68**, 8976-85.
160. Pichler, K., *et al.* (2008) MicroRNA miR-146a and further oncogenesis-related cellular microRNAs are dysregulated in HTLV-1-transformed T lymphocytes. *Retrovirology*, **5**, 100.
161. Matsusaka, K., *et al.* (2011) Classification of Epstein-Barr virus-positive gastric cancers by definition of DNA methylation epigenotypes. *Cancer Res*, **71**, 7187-97.
162. Paschos, K., *et al.* (2009) Epstein-barr virus latency in B cells leads to epigenetic repression and CpG methylation of the tumour suppressor gene Bim. *PLoS Pathog*, **5**, e1000492.
163. Hino, R., *et al.* (2009) Activation of DNA methyltransferase 1 by EBV latent membrane protein 2A leads to promoter hypermethylation of PTEN gene in gastric carcinoma. *Cancer Res*, **69**, 2766-74.
164. Tsai, C.L., *et al.* (2006) Activation of DNA methyltransferase 1 by EBV LMP1 Involves c-Jun NH(2)-terminal kinase signaling. *Cancer Res*, **66**, 11668-76.
165. Ushiku, T., *et al.* (2007) p73 gene promoter methylation in Epstein-Barr virus-associated gastric carcinoma. *Int J Cancer*, **120**, 60-6.
166. Chang, M.S., *et al.* (2006) CpG island methylation status in gastric carcinoma with and without infection of Epstein-Barr virus. *Clin Cancer Res*, **12**, 2995-3002.
167. Kim, J., *et al.* (2005) Silencing and CpG island methylation of GSTP1 is rare in ordinary gastric carcinomas but common in Epstein-Barr virus-associated gastric carcinomas. *Anticancer Res*, **25**, 4013-9.
168. Sakuma, K., *et al.* (2004) High-density methylation of p14ARF and p16INK4A in Epstein-Barr virus-associated gastric carcinoma. *Int J Cancer*, **112**, 273-8.
169. Sudo, M., *et al.* (2004) Promoter hypermethylation of E-cadherin and its abnormal expression in Epstein-Barr virus-associated gastric carcinoma. *Int J Cancer*, **109**, 194-9.
170. Kang, G.H., *et al.* (2002) Epstein-barr virus-positive gastric carcinoma demonstrates frequent aberrant methylation of multiple genes and constitutes CpG island methylator phenotype-positive gastric carcinoma. *Am J Pathol*, **160**, 787-94.
171. Skalska, L., *et al.* (2010) Epigenetic repression of p16(INK4A) by latent Epstein-Barr virus requires the interaction of EBNA3A and EBNA3C with CtBP. *PLoS Pathog*, **6**, e1000951.
172. Grafodatskaya, D., *et al.* (2010) EBV transformation and cell culturing destabilizes DNA methylation in human lymphoblastoid cell lines. *Genomics*, **95**, 73-83.
173. Seo, S.Y., *et al.* (2008) Epstein-Barr virus latent membrane protein 1 suppresses the growth-inhibitory effect of retinoic acid by inhibiting retinoic acid receptor-beta2 expression via DNA methylation. *Cancer Lett*, **270**, 66-76.

For Peer Review

Downloaded from http://carcin.oxfordjournals.org/ at Imperial College London Library on June 15, 2013

174. Maruo, S., *et al.* (2011) Epstein-Barr virus nuclear antigens 3C and 3A maintain lymphoblastoid cell growth by repressing p16INK4A and p14ARF expression. *Proc Natl Acad Sci U S A*, **108**, 1919-24.
175. Anderton, J.A., *et al.* (2011) The H3K27me3 demethylase, KDM6B, is induced by Epstein-Barr virus and over-expressed in Hodgkin's Lymphoma. *Oncogene*, **30**, 2037-43.
176. Knight, J.S., *et al.* (2003) Epstein-Barr virus nuclear antigen 3C recruits histone deacetylase activity and associates with the corepressors mSin3A and NCoR in human B-cell lines. *J Virol*, **77**, 4261-72.
177. Cotter, M.A., 2nd, *et al.* (2000) Modulation of histone acetyltransferase activity through interaction of epstein-barr nuclear antigen 3C with prothymosin alpha. *Mol Cell Biol*, **20**, 5722-35.
178. Radkov, S.A., *et al.* (1999) Epstein-Barr virus nuclear antigen 3C interacts with histone deacetylase to repress transcription. *J Virol*, **73**, 5688-97.
179. Shinozaki, A., *et al.* (2010) Downregulation of microRNA-200 in EBV-associated gastric carcinoma. *Cancer Res*, **70**, 4719-27.
180. Cosmopoulos, K., *et al.* (2009) Comprehensive profiling of Epstein-Barr virus microRNAs in nasopharyngeal carcinoma. *J Virol*, **83**, 2357-67.
181. Kim do, N., *et al.* (2007) Expression of viral microRNAs in Epstein-Barr virus-associated gastric carcinoma. *J Virol*, **81**, 1033-6.
182. Yu, H., *et al.* (2012) Epstein-Barr virus downregulates microRNA 203 through the oncoprotein latent membrane protein 1: a contribution to increased tumor incidence in epithelial cells. *J Virol*, **86**, 3088-99.
183. Feederle, R., *et al.* (2011) A viral microRNA cluster strongly potentiates the transforming properties of a human herpesvirus. *PLoS Pathog*, **7**, e1001294.
184. Du, Z.M., *et al.* (2011) Upregulation of MiR-155 in nasopharyngeal carcinoma is partly driven by LMP1 and LMP2A and downregulates a negative prognostic marker JMJD1A. *PLoS One*, **6**, e19137.
185. Imig, J., *et al.* (2011) microRNA profiling in Epstein-Barr virus-associated B-cell lymphoma. *Nucleic Acids Res*, **39**, 1880-93.
186. Anastasiadou, E., *et al.* (2010) Epstein-Barr virus encoded LMP1 downregulates TCL1 oncogene through miR-29b. *Oncogene*, **29**, 1316-28.
187. Godshalk, S.E., *et al.* (2008) Epstein-Barr virus-mediated dysregulation of human microRNA expression. *Cell Cycle*, **7**, 3595-600.
188. Cameron, J.E., *et al.* (2008) Epstein-Barr virus growth/latency III program alters cellular microRNA expression. *Virology*, **382**, 257-66.
189. Gatto, G., *et al.* (2008) Epstein-Barr virus latent membrane protein 1 trans-activates miR-155 transcription through the NF-kappaB pathway. *Nucleic Acids Res*, **36**, 6608-19.
190. Rahadiani, N., *et al.* (2008) Latent membrane protein-1 of Epstein-Barr virus induces the expression of B-cell integration cluster, a precursor form of microRNA-155, in B lymphoma cell lines. *Biochem Biophys Res Commun*, **377**, 579-83.
191. Motsch, N., *et al.* (2007) Epstein-Barr virus-encoded latent membrane protein 1 (LMP1) induces the expression of the cellular microRNA miR-146a. *RNA Biol*, **4**, 131-7.
192. Lu, F., *et al.* (2008) Epstein-Barr virus-induced miR-155 attenuates NF-kappaB signaling and stabilizes latent virus persistence. *J Virol*, **82**, 10436-43.
193. Guo, W., *et al.* (2011) Alcohol exposure decreases CREB binding protein expression and histone acetylation in the developing cerebellum. *PLoS One*, **6**, e19351.
194. Guo, X.B., *et al.* (2011) Helicobacter pylori induces promoter hypermethylation and downregulates gene expression of IRX1 transcription factor on human gastric mucosa. *J Gastroenterol Hepatol*, **26**, 1685-90.

For Peer Review

Downloaded from http://carcin.oxfordjournals.org/ at Imperial College London Library on June 15, 2013

195. Yan, J., *et al.* (2011) Helicobacter pylori infection promotes methylation of WWOX gene in human gastric cancer. *Biochem Biophys Res Commun*, **408**, 99-102.
196. Yoshida, T., *et al.* (2011) Alu and Satalpha hypomethylation in Helicobacter pylori-infected gastric mucosae. *Int J Cancer*, **128**, 33-9.
197. Shin, C.M., *et al.* (2010) Role of Helicobacter pylori infection in aberrant DNA methylation along multistep gastric carcinogenesis. *Cancer Sci*, **101**, 1337-46.
198. Niwa, T., *et al.* (2010) Inflammatory processes triggered by Helicobacter pylori infection cause aberrant DNA methylation in gastric epithelial cells. *Cancer Res*, **70**, 1430-40.
199. Sepulveda, A.R., *et al.* (2010) CpG methylation and reduced expression of O6-methylguanine DNA methyltransferase is associated with Helicobacter pylori infection. *Gastroenterology*, **138**, 1836-44.
200. Kondo, T., *et al.* (2009) Accumulation of aberrant CpG hypermethylation by Helicobacter pylori infection promotes development and progression of gastric MALT lymphoma. *Int J Oncol*, **35**, 547-57.
201. Nakajima, T., *et al.* (2009) The presence of a methylation fingerprint of Helicobacter pylori infection in human gastric mucosae. *Int J Cancer*, **124**, 905-10.
202. Dong, C.X., *et al.* (2009) Promoter methylation of p16 associated with Helicobacter pylori infection in precancerous gastric lesions: a population-based study. *Int J Cancer*, **124**, 434-9.
203. Leung, W.K., *et al.* (2006) Effects of Helicobacter pylori eradication on methylation status of E-cadherin gene in noncancerous stomach. *Clin Cancer Res*, **12**, 3216-21.
204. Chan, A.O., *et al.* (2003) Promoter methylation of E-cadherin gene in gastric mucosa associated with Helicobacter pylori infection and in gastric cancer. *Gut*, **52**, 502-6.
205. Peterson, A.J., *et al.* (2010) Helicobacter pylori infection promotes methylation and silencing of trefoil factor 2, leading to gastric tumor development in mice and humans. *Gastroenterology*, **139**, 2005-17.
206. Pero, R., *et al.* (2011) Chromatin and DNA methylation dynamics of Helicobacter pylori-induced COX-2 activation. *Int J Med Microbiol*, **301**, 140-9.
207. Bussiere, F.I., *et al.* (2010) H. pylori-induced promoter hypermethylation downregulates USF1 and USF2 transcription factor gene expression. *Cell Microbiol*, **12**, 1124-33.
208. Katayama, Y., *et al.* (2009) Helicobacter pylori causes runx3 gene methylation and its loss of expression in gastric epithelial cells, which is mediated by nitric oxide produced by macrophages. *Biochem Biophys Res Commun*, **388**, 496-500.
209. Ding, S.Z., *et al.* (2010) Helicobacter pylori-induced histone modification, associated gene expression in gastric epithelial cells, and its implication in pathogenesis. *PLoS One*, **5**, e9875.
210. Fehri, L.F., *et al.* (2009) Helicobacter pylori-induced modification of the histone H3 phosphorylation status in gastric epithelial cells reflects its impact on cell cycle regulation. *Epigenetics*, **4**, 577-86.
211. Xia, G., *et al.* (2008) Helicobacter pylori regulates p21(WAF1) by histone H4 acetylation. *Biochem Biophys Res Commun*, **369**, 526-31.
212. Angrisano, T., *et al.* (2012) Helicobacter pylori regulates iNOS promoter by histone modifications in human gastric epithelial cells. *Med Microbiol Immunol*, **201**, 249-57.
213. Matsushima, K., *et al.* (2011) MicroRNA signatures in Helicobacter pylori-infected gastric mucosa. *Int J Cancer*, **128**, 361-70.
214. Belinsky, S.A., *et al.* (2002) Aberrant CpG island methylation of the p16(INK4a) and estrogen receptor genes in rat lung tumors induced by particulate carcinogens. *Carcinogenesis*, **23**, 335-9.
215. Christensen, B.C., *et al.* (2008) Asbestos exposure predicts cell cycle control gene promoter methylation in pleural mesothelioma. *Carcinogenesis*, **29**, 1555-9.

For Peer Review

Downloaded from http://carcin.oxfordjournals.org/ at Imperial College London Library on June 15, 2013

216. Nymark, P., *et al.* (2011) Integrative analysis of microRNA, mRNA and aCGH data reveals asbestos- and histology-related changes in lung cancer. *Genes Chromosomes Cancer*, **50**, 585-97.
217. Krakowczyk, L., *et al.* (2010) Effects of X-ray irradiation on methylation levels of p16, MGMT and APC genes in sporadic colorectal carcinoma and corresponding normal colonic mucosa. *Med Sci Monit*, **16**, CR469-74.
218. Giotopoulos, G., *et al.* (2006) DNA methylation during mouse hemopoietic differentiation and radiation-induced leukemia. *Exp Hematol*, **34**, 1462-70.
219. Cleary, H.J., *et al.* (1999) Allelic loss and promoter hypermethylation of the p15INK4b gene features in mouse radiation-induced lymphoid - but not myeloid - leukaemias. *Leukemia*, **13**, 2049-52.
220. Koturbash, I., *et al.* (2007) Role of epigenetic effectors in maintenance of the long-term persistent bystander effect in spleen in vivo. *Carcinogenesis*, **28**, 1831-8.
221. Templin, T., *et al.* (2011) Radiation-induced micro-RNA expression changes in peripheral blood cells of radiotherapy patients. *Int J Radiat Oncol Biol Phys*, **80**, 549-57.
222. Ding, N., *et al.* (2011) Detection of novel human MiRNAs responding to X-ray irradiation. *J Radiat Res*, **52**, 425-32.
223. Kumar, A., *et al.* (2011) gamma-radiation induces cellular sensitivity and aberrant methylation in human tumor cell lines. *Int J Radiat Biol*, **87**, 1086-96.
224. Cui, W., *et al.* (2011) Plasma miRNA as biomarkers for assessment of total-body radiation exposure dosimetry. *PLoS One*, **6**, e22988.
225. Templin, T., *et al.* (2011) Whole mouse blood microRNA as biomarkers for exposure to gamma-rays and (56)Fe ion. *Int J Radiat Biol*, **87**, 653-62.
226. Liu, Y., *et al.* (2008) Aberrant gene promoter methylation in sputum from individuals exposed to smoky coal emissions. *Anticancer Res*, **28**, 2061-6.
227. Huang, Y., *et al.* (2011) Cigarette smoke induces promoter methylation of single-stranded DNA-binding protein 2 in human esophageal squamous cell carcinoma. *Int J Cancer*, **128**, 2261-73.
228. Lin, R.K., *et al.* (2010) The tobacco-specific carcinogen NNK induces DNA methyltransferase 1 accumulation and tumor suppressor gene hypermethylation in mice and lung cancer patients. *J Clin Invest*, **120**, 521-32.
229. Oka, D., *et al.* (2009) The presence of aberrant DNA methylation in noncancerous esophageal mucosae in association with smoking history: a target for risk diagnosis and prevention of esophageal cancers. *Cancer*, **115**, 3412-26.
230. Marsit, C.J., *et al.* (2006) Carcinogen exposure and gene promoter hypermethylation in bladder cancer. *Carcinogenesis*, **27**, 112-6.
231. Marsit, C.J., *et al.* (2006) Epigenetic inactivation of the SFRP genes is associated with drinking, smoking and HPV in head and neck squamous cell carcinoma. *Int J Cancer*, **119**, 1761-6.
232. Kikuchi, S., *et al.* (2006) Hypermethylation of the TSLC1/IGSF4 promoter is associated with tobacco smoking and a poor prognosis in primary nonsmall cell lung carcinoma. *Cancer*, **106**, 1751-8.
233. Liu, Y., *et al.* (2006) Aberrant promoter methylation of p16 and MGMT genes in lung tumors from smoking and never-smoking lung cancer patients. *Neoplasia*, **8**, 46-51.
234. Lea, J.S., *et al.* (2004) Aberrant p16 methylation is a biomarker for tobacco exposure in cervical squamous cell carcinogenesis. *Am J Obstet Gynecol*, **190**, 674-9.
235. Anttila, S., *et al.* (2003) Methylation of cytochrome P4501A1 promoter in the lung is associated with tobacco smoking. *Cancer Res*, **63**, 8623-8.
236. Kim, D.H., *et al.* (2001) p16(INK4a) and histology-specific methylation of CpG islands by exposure to tobacco smoke in non-small cell lung cancer. *Cancer Res*, **61**, 3419-24.

For Peer Review

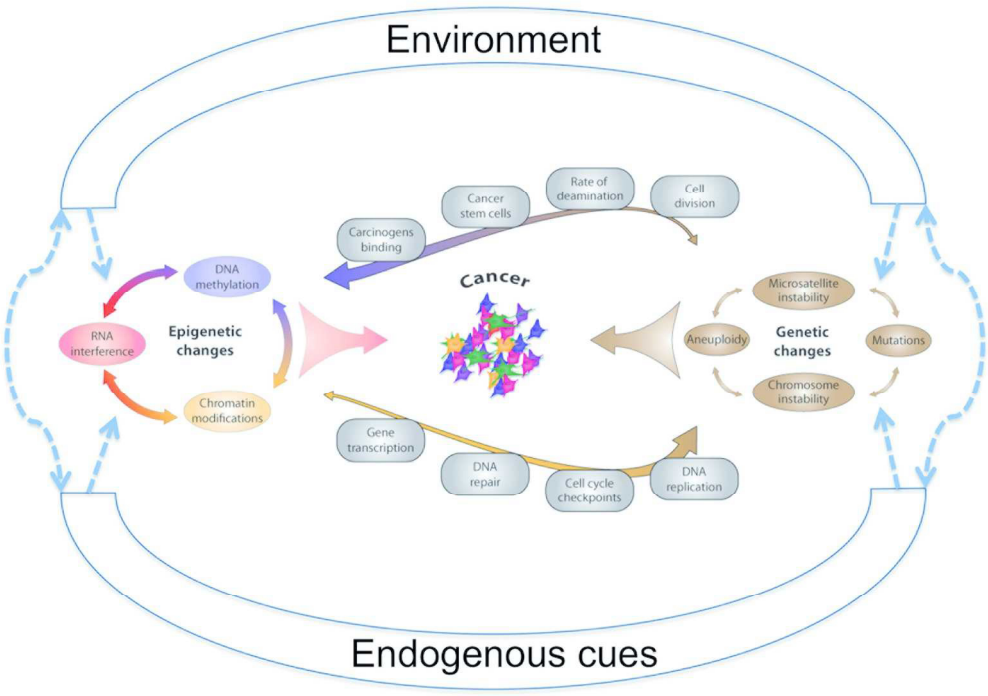
Downloaded from http://carcin.oxfordjournals.org/ at Imperial College London Library on June 15, 2013

237. Hussain, M., *et al.* (2009) Tobacco smoke induces polycomb-mediated repression of Dickkopf-1 in lung cancer cells. *Cancer Res*, **69**, 3570-8.
238. Vaissiere, T., *et al.* (2009) Quantitative analysis of DNA methylation profiles in lung cancer identifies aberrant DNA methylation of specific genes and its association with gender and cancer risk factors. *Cancer Res*, **69**, 243-52.
239. Liu, H., *et al.* (2007) Cigarette smoke induces demethylation of prometastatic oncogene synuclein-gamma in lung cancer cells by downregulation of DNMT3B. *Oncogene*, **26**, 5900-10.
240. Adenuga, D., *et al.* (2009) Histone deacetylase 2 is phosphorylated, ubiquitinated, and degraded by cigarette smoke. *Am J Respir Cell Mol Biol*, **40**, 464-73.
241. Schernhammer, E.S., *et al.* (2010) Dietary folate, alcohol and B vitamins in relation to LINE-1 hypomethylation in colon cancer. *Gut*, **59**, 794-9.
242. Figueiredo, J.C., *et al.* (2009) Global DNA hypomethylation (LINE-1) in the normal colon and lifestyle characteristics and dietary and genetic factors. *Cancer Epidemiol Biomarkers Prev*, **18**, 1041-9.
243. Bonsch, D., *et al.* (2006) Lowered DNA methyltransferase (DNMT-3b) mRNA expression is associated with genomic DNA hypermethylation in patients with chronic alcoholism. *J Neural Transm*, **113**, 1299-304.
244. Nan, H.M., *et al.* (2005) Effects of dietary intake and genetic factors on hypermethylation of the hMLH1 gene promoter in gastric cancer. *World J Gastroenterol*, **11**, 3834-41.
245. D'Addario, C., *et al.* (2011) Ethanol and acetaldehyde exposure induces specific epigenetic modifications in the prodynorphin gene promoter in a human neuroblastoma cell line. *FASEB J*, **25**, 1069-75.
246. Hicks, S.D., *et al.* (2010) Ethanol-induced methylation of cell cycle genes in neural stem cells. *J Neurochem*, **114**, 1767-80.
247. Choi, S.W., *et al.* (1999) Chronic alcohol consumption induces genomic but not p53-specific DNA hypomethylation in rat colon. *J Nutr*, **129**, 1945-50.
248. Liu, Y., *et al.* (2009) Alcohol exposure alters DNA methylation profiles in mouse embryos at early neurulation. *Epigenetics*, **4**, 500-11.
249. Hu, Y., *et al.* (2009) Ethanol disrupts chondrification of the neurocranial cartilages in medaka embryos without affecting aldehyde dehydrogenase 1A2 (Aldh1A2) promoter methylation. *Comp Biochem Physiol C Toxicol Pharmacol*, **150**, 495-502.
250. Agudelo, M., *et al.* (2011) Effects of alcohol on histone deacetylase 2 (HDAC2) and the neuroprotective role of trichostatin A (TSA). *Alcohol Clin Exp Res*, **35**, 1550-6.
251. Zhao, Y., *et al.* (2011) Zinc deprivation mediates alcohol-induced hepatocyte IL-8 analog expression in rodents via an epigenetic mechanism. *Am J Pathol*, **179**, 693-702.
252. Avissar, M., *et al.* (2009) MicroRNA expression in head and neck cancer associates with alcohol consumption and survival. *Carcinogenesis*, **30**, 2059-63.
253. Tang, Y., *et al.* (2008) Effect of alcohol on miR-212 expression in intestinal epithelial cells and its potential role in alcoholic liver disease. *Alcohol Clin Exp Res*, **32**, 355-64.

For Peer Review

Downloaded from http://carcin.oxfordjournals.org/ at Imperial College London Library on June 15, 2013

Figure 1

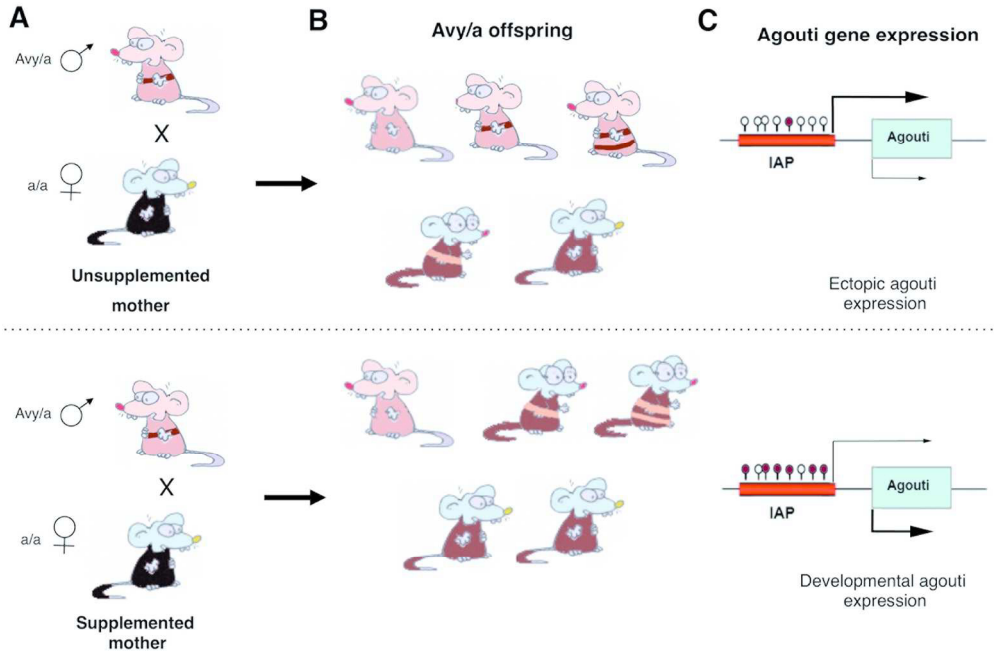


193x156mm (300 x 300 DPI)

For Review

Downloaded from http://carcin.oxfordjournals.org/ at Imperial College London Library on June 15, 2013

Figure 2



173x126mm (300 x 300 DPI)

For Review

Downloaded from http://carcin.oxfordjournals.org/ at Imperial College London Library on June 15, 2013

APPENDIX F

Methylome Analysis and Epigenetic Changes Associated with Menarcheal Age

Methylome Analysis and Epigenetic Changes Associated with Menarcheal Age

Christiana A. Demetriou[1,2], Jia Chen[3], Silvia Polidoro[4], Karin VanVeldhoven[2], Cyrille Cuenin[5], Gianluca Campanella[2], Kevin Brennan[6], Françoise Clavel-Chapelon[7,8], Laure Dossus[7,8], Marina Kvaskoff [7,8], Dagmar Drogan[9], Heiner Boeing[9], Rudolf Kaaks[10], Angela Risch[11], Dimitrios Trichopoulos[12,13,14], Pagona Lagiou[15,16,17] , Giovanna Masala[18], Sabina Sieri[19] , Rosario Tumino[20], Salvatore Panico[21], J Ramón Quirós[22], María-José Sánchez[23,24], Pilar Amiano[24,25], José María Huerta Castaño[24,26], Eva Ardanaz[24,27], Charlotte Onland-Moret[28], Petra Peeters[28], Kay-Tee Khaw[29], Nick Wareham[30], Timothy J. Key[31], Ruth C Travis[31], Isabelle Romieu[32], Valentina Gallo[2,33], Marc Gunter[2], Zdenko Herceg[4], Kyriacos Kyriacou[1], Elio Riboli[34], James M. Flanagan[6], Paolo Vineis[2]

[1] Department of electron Microscopy & Molecular Pathology, The Cyprus Institute of Neurology and Genetics, Nicosia Cyprus

[2] Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK

[3] Departments of Preventive Medicine, Paediatrics, and Oncological Sciences at Mount Sinai School of Medicine, New York, USA

[4] Molecular and Genetic Epidemiology Unit, Human Genetics Foundation, Torino, Italy

[5] Epigenetics Group, International Agency for Research on Cancer, Lyon, France

[6] Epigenetics Unit, Department of Surgery and Cancer, Imperial College London, UK

[7] Inserm, Centre for Research in Epidemiology and Population Health, U1018, Institut Gustave Roussy, F-94805, Villejuif, France.

[8] Paris South University, UMRS 1018, F-94805, Villejuif, France

[9] German Institute of Human Nutrition Potsdam-Rehbruecke, Department of Epidemiology, Arthur-Scheunert-Allee 114–116, D-14558 Nuthetal, Germany

[10] Department of Cancer Epidemiology; German Cancer Research Center [DKFZ], Heidelberg, Germany

[11] Department of Epigenomics and Cancer risk Factors; German Cancer Research Center [DKFZ], Heidelberg, Germany

[12] Department of Epidemiology, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115, USA

[13] Bureau of Epidemiologic Research, Academy of Athens, 28 Panepistimiou Street, Athens, GR-106 79, Greece

[14] Hellenic Health Foundation, 13 Kaisareias Street, Athens, GR-115 27, Greece

[15] WHO Collaborating Center for Food and Nutrition Policies, Department of Hygiene, Epidemiology and Medical Statistics, University of Athens Medical School, 75 M. Asias Street, Goudi, GR-115 27, Athens, Greece

[16] Department of Epidemiology, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115, USA

[17] Bureau of Epidemiologic Research, Academy of Athens, 28 Panepistimiou Street, Athens, GR-106 79, Greece

[18] Molecular and Nutritional Epidemiology Unit, Cancer Research and Prevention Institute – ISPO, Florence, Italy

[19] Epidemiology and Prevention Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Milano, Italy

[20] Cancer Registry and Histopathology Unit, "Civile - M.P. Arezzo" Hospital, ASP Ragusa, Italy

[21] Department of Clinical and Experimental Medicine, Federico II University, Naples, Italy

[22] Public Health Directorate, Asturias, Spain

[23]Andalusian School of Public Health, Granada, Spain
[24] CIBER Epidemiología y Salud Pública (CIBERESP), Madrid, Spain
[25] Public Helath division of Gipuzkoa, BioDonostia Research Institute, Health Department of Basque Region, San Sebastia, Spain
[26] Department of Epidemiology, Murcia Regional Health Council, Murcia, Spain
[27] Navarre Public Health Institute, Pamplona, Spain
[28]Julius Center for Health Sciences and Primary Care University Medical Center, Utrecht, the Netherlands
[29]MRC Epidemiology Unit, Cambridge Institute of Public Health, Cambridge, UK
[30] Cambridge Institute of Public Health, Cambridge, UK
[31] Cancer Epidemiology Unit, Nuffield Department of Clinical Medicine, University of Oxford, Oxford, OX3 0NR, UK
[32]Nutrition and Metabolism Section, International Agency for Research on Cancer, Lyon, France
[33]Centre for Primary Care and Public Health, Barts and the London School of Medicine, Queen Mary, University of London, London, UK
[34]School of Public Health, Imperial College London, London, UK

Short title: Age at menarche and DNA methylation

Key Words: Age at Menarche, reproductive risk factors, DNA Methylation, Breast Cancer, LUminometric Methylation Assay (LUMA), Infinium HumanMethylation 450 BeadChip

Financial Disclosure:
CAD received funding from EU-Europeaid grant CRIS 2009/223-507.The EPIC cohort is supported by the Europe Against Cancer Program of the European Commission (SANCO). The individual centres also received funding from: Denmark: Danish Cancer Society; France: Ligue centre le Cancer, Institut Gustave Roussy, Mutuelle Générale de l'Education Nationale, Institut National de la Santé et de la Recherche Médicale (INSERM); Greece: Hellenic Ministry of Health, the Stavros Niarchos Foundation and the Hellenic Health Foundation; Germany: German Cancer Aid, German Cancer Research Center, and Federal Ministry of Education and Research (Grant 01-EA-9401) ; Italy: Italian Association for Research on Cancer and the National Research Council; The Netherlands: Dutch Ministry of Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund (WCRF); Spain: Health Research Fund (FIS) of the Spanish Ministry of Health (Exp 96/0032) and the participating regional governments and institutions; Sweden: Swedish Cancer Society, Swedish Scientific Council, and Regional Government of Skane; UK: Cancer Research UK and Medical Research Council UK and Breast Cancer Campaign.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflict of Interest:

The authors have no conflict of interest to disclose, financial or of any other form.

Corresponding Author:

Christiana Demetriou

e-mail: christiana.demetriou08@imperial.ac.uk

telephone: (+357) 96553080

Abstract

Reproductive factors have been linked to both breast cancer and DNA methylation, suggesting methylation as an important mechanism by which reproductive factors impact on disease risk. However, few studies have investigated the link between reproductive factors and DNA methylation in humans.

Genome-wide methylation in peripheral blood lymphocytes of 376 healthy women from the prospective EPIC study was investigated using LUminometric Methylation Assay (LUMA). Also, methylation of 458877 CpG sites was additionally investigated in an independent group of 332 participants of the EPIC-Italy sub-cohort, using the Infinium HumanMethylation 450 BeadChip. Multivariate logistic regression and linear models were used to investigate the association between reproductive risk factors and genome wide and CpG-specific DNA methylation, respectively.

Menarcheal age was inversely associated with global DNA methylation as measured with LUMA. For each yearly increase in age at menarche, the risk of having genome wide methylation below median level was increased by 32% (OR:1.32, 95%CI:1.14-1.53). When age at menarche was treated as a categorical variable, there was an inverse dose-response relationship with LUMA methylation levels ($OR_{12\text{-}14vs.\leq 11yrs}$:1.78, 95%CI:1.01-3.17 and $OR_{\geq 15vs.\leq 11yrs}$:4.59, 95%CI:2.04-10.33; P for trend<0.0001). However, average levels of global methylation as measured by the Illumina technology were not significantly associated with menarcheal age. In locus by locus comparative analyses, only one CpG site had significantly different methylation depending on the menarcheal age category examined, but this finding was not replicated by pyrosequencing in an independent data set.

This study suggests a link between age at menarche and genome wide DNA methylation, and the difference in results between the two arrays suggests that repetitive element methylation has a role in the association. Epigenetic changes may be modulated by menarcheal age, or the association may be a mirror of other important changes in early life that have a detectable effect on both methylation levels and menarcheal age.

INTRODUCTION

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

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

Reproductive factors were also shown to impact on global DNA methylation. Terry et al. [18] showed that factors that impact on breast cancer risk, including a greater birth height, a later age at menarche, nulliparity, and a later age at first birth were associated with higher global DNA methylation levels, but these results were

not replicated in other studies [8,24]. However, the studies that investigate reproductive factors and epigenetic alterations are few [25].

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

RESULTS

LUMA methylation is associated with menarcheal age

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

The association between age at menarche and methylation was further examined using logistic regression to adjust for potential confounders. When median methylation was used as a cut off, 194 subjects had methylation below median levels and 182 had methylation levels above median. Using these two classes, logistic regression showed that age at menarche was significantly associated with class occupancy. As shown in Table 3, for every yearly increase in age at menarche, the risk of having below median methylation was increased by 32% (OR: 1.32, 95%CI: 1.14-1.53). When age at menarche was treated as a categorical variable, there was an inverse dose-response relationship with methylation levels: for the age category 12-14 compared to ≤11years the OR was 1.78 (95% CI: 1.01- 3.17), and for the age category ≥15 compared to ≤11years the OR was 4.59 (95% CI: 2.04-10.33) (P for trend<0.0001). These significant associations persisted even after adjustment for relevant confounders: centre, plate number, age at blood collection, height, weight, total physical activity, smoking status, daily alcohol consumption, and daily folate consumption.

Illumina 450k methylome analysis identifies an epi-allele associated with menarcheal age

In the second population group, 330 subjects (out of 332) had available information on age at menarche (Table 1). In contrast to LUMA global methylation, the median genome-wide methylation level using the 450k ILLUMINA assay did not significantly differ between menarcheal age groups. Similarly, CpG island methylation and promoter methylation were not significantly different between subjects in different menarcheal age categories. However, there was a trend towards decreasing methylation with increasing age at menarche, consistent with the LUMA results (Figure 1).

When adjusting for case-control status, age, and position on the chip in a linear regression model with methylation M-values as a continuous outcome, and age at menarche as a categorical variable (>11yrs vs. ≤11yrs), age at menarche was significantly associated with methylation in a single CpG site (cg01339004), located on the body of the SMAD6 gene ($p<1.00x10^{-7}$, genome-wide level significance) (Table 4, Figure 2). When only those subjects that remained healthy for at least 5 years following recruitment and blood collection were analysed, the same CpG site was found to be significantly associated with age at menarche ($p=6.71x10^{-8}$).

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

DISCUSSION

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

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

Given the conflicting reports in the literature[8,15,18], we aimed to replicate, in a dataset with whole genome methylation data, the association between age at menarche and DNA methylation that we observed with the LUMA technology. This was done by using the robust Illumina technology. This approach also enabled the identification of specific genes that might be involved in the mechanistic pathways linking menarcheal age with disease. In contrast to the findings of LUMA, genome wide methylation in this second dataset did not significantly differ between subjects in different menarcheal age groups. However, there was a non-significant trend towards decreasing methylation with increasing age at menarche, consistent with the LUMA findings (Figure 1). The lack of association in this dataset could be caused by differences in coverage between the two assays used. LUMA assesses methylation of a specific restriction enzyme site (HpaII, CCGG), which occurs most frequently in CpG island promoters – also covered by the 450K array – but also in repetitive elements. However, the Infinium HumanMethylation 450 BeadChip, due to its probe design, does not interrogate repetitive elements which are found to be differentially methylated in many cases of neoplasia [28,29]. For example, satellite and SINE repeats were found to be enriched with hypomethylated Differentially Methylated

Regions (DMRs) whereas LINE was enriched with hypermethylated DMRs in malignant peripheral nerve sheath tumours compared to normal Schwann cells [30]. If age at menarche is related to methylation patterns in these repetitive elements, the hypomethylation would not have been evident in the 450K chip but it would have been detected in the LUMA assay. This suggests that the LUMA based association is being driven largely by methylation differences in repetitive elements, where age at menarche could have a greater effect.

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

The mechanistic link which could explain the association between menarcheal age and genome-wide DNA methylation, but not in individual CpG loci is yet to be determined. However, endogenous oestrogen exposure is a strong candidate for epigenetic changes since an earlier age at menarche exposes a woman to a greater cumulative amount of endogenous oestrogens over her lifetime, and there have been reports which show that oestrogen impacts on DNA methylation. More specifically it was shown that oestrogen receptor (ER) positive breast tumour tissues have differential methylation at several CpG loci compared to ER negative tumours [24,31] and oestrogen induced breast tumours have differential DNA methylation patterns in ACI rat mammary gland tissue [32]. Further investigation into the role of oestrogen on repetitive element methylation is, therefore, warranted.

It is also possible that age at menarche is an indirect indicator of other macroscopic changes that may impact on DNA methylation. It is widely accepted that development is plastic and the sensitivity of the epigenetic system to environmental factors is heightened during periods of developmental plasticity such as childhood, adolescence and puberty [17]. Epigenetic modifications in response to environmental exposures at these critical periods are often subtle initially and even though they do not lead to phenotypic changes at the time of exposure, they may lead to increased risk of dysfunction and disease later on in life [17]. Trends in the past decades show a rapid shift towards an earlier age at menarche and this is more pronounced in developed countries [33]. This trend is too steep to be attributed to genetic changes. Instead, environmental exposures at the periods of developmental plasticity are likely to be the cause of the dramatic decrease in age at menarche. For example, childhood obesity disrupts the hormonal milieu leading to an increase in adipocyte secreted

leptin, or in adrenal secreted androgens, all of which impact on menarcheal onset [34]. Pre or neo-natal nutrition as well as early life exposure to endocrine disrupting chemicals (EDCs) can also lead to hormonal imbalances impacting on age at menarche [17]. Given that these exposures occur at the periods when the epigenetic signature is more plastic, they might also lead to aberrant DNA methylation changes which will be inherited during cell divisions and be detectable years later. Therefore, the aberrant DNA methylation pattern observed in adulthood might not be related to menarcheal age per se but to an early life environmental exposure, like diet, that impacts both on age at menarche and on DNA methylation.

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

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

MATERIALS and METHODS

Stage 1: genome wide methylation with LUminometric Methylation Assay (LUMA)

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

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

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

$$\text{Genome wide Methylation (\%)} = [1-(\text{HpaII}\Sigma G/\Sigma T)/(\text{MspI}\Sigma G/\Sigma T)*100 \text{ [37].}$$

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

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

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

Stage 2: locus-by-locus analysis with Illumina 450K to replicate the findings of LUMA genome wide methylation analysis

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

The EPIC Italy sub-cohort consists of 32,578 female subjects recruited from 5 different centers – Varese, Turin, Florence, Naples, and Ragusa. From this subcohort, 166 breast cancer cases and 166 controls, matched on date of birth (± 5 years), seasonality of blood draw, and date of recruitment were selected. However, since for this investigation case/control status is not the outcome and since at the time of blood collection all individuals were healthy, all 332 blood samples were treated as healthy blood. Nevertheless, given the long latency period of neoplasia, analyses were carried out on all subjects with age at menarche information (n=330) as well as on only the subjects that remained healthy at least 5 years following recruitment and blood collection (n=240).

Illumina 450K methylation: DNA was extracted from buffy coats or blood cell fractions using the QIAsymphony DNA Midi Kit (Qiagen, Crawley, UK). 500 ng of DNA was bisulphite-converted with the EZ-96 DNA Methylation-Gold™ Kit, used according to the manufacturer's protocol (Zymo Research, Orange, CA, USA). Next, the 450K DNA methylation array by Illumina (Infinium HumanMethylation 450 BeadChip) was performed on 4 µl of bisulphite-converted DNA, following the Illumina Infinium HD Methylation protocol. This array includes 485,577 cytosine positions of the human genome (482,421 CpG sites (99.4%), 3091 non-CpG sites and 65 random SNPs; hereafter the term CpG will be used to refer to all of these,

unless otherwise specified). Briefly, a whole genome amplification step was followed by enzymatic end-point fragmentation and hybridization to HumanMethylation 450 BeadChips at 48°C for 17 h, followed by single nucleotide extension. The incorporated nucleotides were labelled with biotin (ddCTP and ddGTP) and 2,4-dinitrophenol (DNP) (ddATP and ddTTP). After the extension step and staining, the BeadChip was washed and scanned using the Illumina HiScan SQ scanner. The intensities of the images were extracted using the GenomeStudio (v.2011.1) Methylation module (1.9.0) software, which normalizes within-sample data using different internal controls that are present on the HumanMethylation 450 BeadChip and internal background probes.

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

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

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

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

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

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

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

Q-values, measuring the maximum False Discovery Rate (FDR) from the Benjamini and Hochberg method were derived for each probe analysis, and overall Type I error was controlled for by conservatively applying Bonferroni multiple testing correction. The per-test significance cut-off value was set to 1.00×10^{-7}. The analysis was performed first on all 332 subjects and then repeated only on the 240 subjects that remained healthy for at least five years following recruitment. The linear model was adjusted for age, case-control status and chip position. All confounders were entered into the model as continuous variables with the exception of chip position.

Stage 3: single locus (SMAD6, cg01339004) pyrosequencing analysis to replicate the findings of locus by locus Illumina 450K analysis

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

Bisulphite Pyrosequencing was used to quantify CpG specific methylation at the SMAD6 locus in these individuals using standard protocols [43]. The primers used for cg01339004 were Forward ([BIOTIN]–

TGGTATAGTAGTGGTTTGGTATAAGAT), Reverse (TACCACCCACCCATTCACTCTATAA) and Sequencing Primer (TCTATAAATAAACAAACTAAAACC).

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

AKNOWLEDGEMENTS

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

References

1. Jones PA, Baylin SB (2007) The epigenomics of cancer. Cell 128: 683–692. doi:10.1016/j.cell.2007.01.029.

2. Jovanovic J, Rønneberg JA, Tost J, Kristensen V (2010) The epigenetics of breast cancer. Mol Oncol 4: 242–254. doi:10.1016/j.molonc.2010.04.002.

3. Esteller M, Corn PG, Baylin SB, Herman JG (2001) A gene hypermethylation profile of human cancer. Cancer Res 61: 3225–3229.

4. Esteller M, Herman JG (2002) Cancer as an epigenetic disease: DNA methylation and chromatin alterations in human tumours. J Pathol 196: 1–7. doi:10.1002/path.1024.

5. Portela A, Esteller M (2010) Epigenetic modifications and human disease. Nat Biotech 28: 1057–1068. doi:10.1038/nbt.1685.

6. Veeck J, Esteller M (2010) Breast cancer epigenetics: from DNA methylation to microRNAs. J Mammary Gland Biol Neoplasia 15: 5–17. doi:10.1007/s10911-010-9165-1.

7. Cho YH, Yazici H, Wu H-C, Terry MB, Gonzalez K, et al. (2010) Aberrant promoter hypermethylation and genomic hypomethylation in tumor, adjacent normal tissues and blood from breast cancer patients. Anticancer Res 30: 2489–2496.

8. Choi J-Y, James SR, Link PA, McCann SE, Hong C-C, et al. (2009) Association between global DNA hypomethylation in leukocytes and risk of breast cancer. Carcinogenesis 30: 1889 –1897. doi:10.1093/carcin/bgp143.

9. Lim U, Flood A, Choi S, Albanes D, Cross AJ, et al. (2008) Genomic Methylation of Leukocyte DNA in Relation to Colorectal Adenoma Among Asymptomatic Women. Gastroenterology 134: 47–55. doi:10.1053/j.gastro.2007.10.013.

10. Pufulete M, Al-Ghnaniem R, Leather AJM, Appleby P, Gout S, et al. (2003) Folate status, genomic DNA hypomethylation, and risk of colorectal adenoma and cancer: a case control study. Gastroenterology 124: 1240–1248.

11. Hou L (2010) Blood leukocyte DNA hypomethylation and gastric cancer risk in a high-risk Polish population. International Journal of Cancer Journal International Du Cancer 127: 1866–1874. doi:10.1002/ijc.25190.

12. Hsiung DT, Marsit CJ, Houseman EA, Eddy K, Furniss CS, et al. (2007) Global DNA methylation level in whole blood as a biomarker in head and neck squamous cell carcinoma. Cancer Epidemiol Biomarkers Prev 16: 108–114. doi:10.1158/1055-9965.EPI-06-0636.

13. Moore LE, Pfeiffer RM, Poscablo C, Real FX, Kogevinas M, et al. (2008) Genomic DNA hypomethylation as a biomarker for bladder cancer susceptibility in the Spanish Bladder Cancer Study: a case-control study. Lancet Oncol 9: 359–366. doi:10.1016/S1470-2045(08)70038-X.

14. Brennan K, Flanagan JM (2012) Is there a link between genome-wide hypomethylation in blood and cancer risk? Cancer Prev Res (Phila) 5: 1345–1357. doi:10.1158/1940-6207.CAPR-12-0316.

15. Xu X, Gammon MD, Hernandez-Vargas H, Herceg Z, Wetmur JG, et al. (2012) DNA methylation in peripheral blood measured by LUMA is associated with breast cancer in a population-based study. FASEB J 26: 2657–2666. doi:10.1096/fj.11-197251.

16. Bird A (2002) DNA methylation patterns and epigenetic memory. Genes Dev 16: 6–21. doi:10.1101/gad.947102.

17. Barouki R, Gluckman PD, Grandjean P, Hanson M, Heindel JJ (2012) Developmental origins of non-communicable disease: implications for research and public health. Environ Health 11: 42. doi:10.1186/1476-069X-11-42.

18. Terry MB, Ferris JS, Pilsner R, Flom JD, Tehranifar P, et al. (2008) Genomic DNA Methylation among Women in a Multiethnic New York City Birth Cohort. Cancer Epidemiology Biomarkers & Prevention 17: 2306 –2310. doi:10.1158/1055-9965.EPI-08-0312.

19. Terry MB, Delgado-Cruzata L, Vin-Raviv N, Wu HC, Santella RM (2011) DNA methylation in white blood cells: association with risk factors in epidemiologic studies. Epigenetics 6: 828–837.

20. Scoccianti C, Ricceri F, Ferrari P, Cuenin C, Sacerdote C, et al. (2011) Methylation patterns in sentinel genes in peripheral blood cells of heavy smokers: Influence of cruciferous vegetables in an intervention study. epigenetics 6: 1114–1119. doi:10.4161/epi.6.9.16515.

21. Herceg Z (2007) Epigenetics and cancer: towards an evaluation of the impact of environmental and dietary factors. Mutagenesis 22: 91 –103. doi:10.1093/mutage/gel068.

22. Zhang FF, Morabia A, Carroll J, Gonzalez K, Fulda K, et al. (2011) Dietary patterns are associated with levels of global genomic DNA methylation in a cancer-free population. J Nutr 141: 1165–1171. doi:10.3945/jn.110.134536.

23. Piyathilake C, Badiga S, Johanning G, Alvarez R, Partridge E (2011) Predictors and Health Consequences of Epigenetic Changes Associated with Excess Body Weight in Women of Child-bearing Age. Cancer Epidemiol Biomarkers Prev 20: 719–719. doi:10.1158/1055-9965.EPI-11-0094.

24. Christensen BC, Kelsey KT, Zheng S, Houseman EA, Marsit CJ, et al. (2010) Breast Cancer DNA Methylation Profiles Are Associated with Tumor Size and Alcohol and Folate Intake. PLoS Genet 6: e1001043. doi:10.1371/journal.pgen.1001043.

25. Teegarden D, Romieu I, Lelièvre SA (2012) Redefining the impact of nutrition on breast cancer incidence: is epigenetics involved? Nutr Res Rev 25: 68–95. doi:10.1017/S0954422411000199.

26. Widschwendter M, Jones PA (2002) DNA methylation and breast carcinogenesis. Oncogene 21: 5462–5482. doi:10.1038/sj.onc.1205606.

27. Karimi M, Johansson S, Stach D, Corcoran M, Grandér D, et al. (2006) LUMA (LUminometric Methylation Assay)--a high throughput method to the analysis of genomic DNA methylation. Exp Cell Res 312: 1989–1995. doi:10.1016/j.yexcr.2006.03.006.

28. Ehrlich M (2002) DNA methylation in cancer: too much, but also too little. Oncogene 21: 5400–5413. doi:10.1038/sj.onc.1205651.

29. Wilson AS, Power BE, Molloy PL (2007) DNA hypomethylation and human diseases. Biochim Biophys Acta 1775: 138–162. doi:10.1016/j.bbcan.2006.08.007.

30. Feber A, Wilson GA, Zhang L, Presneau N, Idowu B, et al. (2011) Comparative methylome analysis of benign and malignant peripheral nerve sheath tumors. Genome Res 21: 515–524. doi:10.1101/gr.109678.110.

31. Flanagan JM, Cocciardi S, Waddell N, Johnstone CN, Marsh A, et al. (2010) DNA methylome of familial breast cancer identifies distinct profiles defined by mutation status. Am J Hum Genet 86: 420–433. doi:10.1016/j.ajhg.2010.02.008.

32. Starlard-Davenport A, Tryndyak VP, James SR, Karpf AR, Latendresse JR, et al. (2010) Mechanisms of epigenetic silencing of the Rassf1a gene during estrogen-induced breast carcinogenesis in ACI rats. Carcinogenesis 31: 376–381. doi:10.1093/carcin/bgp304.

33. Tanner JM (1973) Trend towards Earlier Menarche in London, Oslo, Copenhagen, the Netherlands and Hungary. , Published online: 11 May 1973; | doi:101038/243095a0 243: 95–96. doi:10.1038/243095a0.

34. Ahmed ML, Ong KK, Dunger DB (2009) Childhood obesity and the timing of puberty. Trends Endocrinol Metab 20: 237–242. doi:10.1016/j.tem.2009.02.004.

35. Riboli E, Hunt KJ, Slimani N, Ferrari P, Norat T, et al. (2002) European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. Public Health Nutr 5: 1113–1124. doi:10.1079/PHN2002394.

36. Karimi M, Johansson S, Ekström TJ (2006) Using LUMA: a Luminometric-based assay for global DNA-methylation. Epigenetics 1: 45–48.

37. Bjornsson HT, Sigurdsson MI, Fallin MD, Irizarry RA, Aspelund T, et al. (2008) Intra-individual change over time in DNA methylation with familial clustering. JAMA 299: 2877–2883. doi:10.1001/jama.299.24.2877.

38. Triche TJ, Weisenberger DJ, Berg DVD, Laird PW, Siegmund KD (2013) Low-level processing of Illumina Infinium DNA Methylation BeadArrays. Nucl Acids Res. Available: http://nar.oxfordjournals.org/content/early/2013/03/09/nar.gkt090. Accessed 11 May 2013.

39. Dedeurwaerder S, Defrance M, Calonne E, Denis H, Sotiriou C, et al. (2011) Evaluation of the Infinium Methylation 450K technology. Epigenomics 3: 771–784. doi:10.2217/epi.11.105.

40. Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, et al. (2010) Tackling the widespread and critical impact of batch effects in high-throughput data. Nature Reviews Genetics 11: 733–739. doi:10.1038/nrg2825.

41. Bock C (2012) Analysing and interpreting DNA methylation data. Nature Reviews Genetics 13: 705–719. doi:10.1038/nrg3273.

42. Du P, Zhang X, Huang C-C, Jafari N, Kibbe WA, et al. (2010) Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformatics 11: 587. doi:10.1186/1471-2105-11-587.

43. Shenker NS, Polidoro S, van Veldhoven K, Sacerdote C, Ricceri F, et al. (2012) Epigenome-wide association study in the European Prospective Investigation into Cancer and Nutrition (EPIC-Turin) identifies novel genetic loci associated with smoking. Hum Mol Genet. doi:10.1093/hmg/dds488.

Figure 1 – Boxplots of median genome-wide methylation between the three menarcheal age categories.
A: Median % global methylation as measured with LUMA in Stage 1. Bi. Genome-wide methylation across all probes (averaged per individual). Bii. Genome-wide methylation across probes on CpG islands (averaged per individual). Biii. Genome-wide methylation across probes on promoter regions (averaged per individual)

$M^{¤}$= Median methylation value

p= p value from Wilcoxon rank-sum test comparisons

Figure 2 – Analysis of SMAD6 cg01339004 probe methylation

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

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

Covariate		Stage 1 (n = 376)*	Stage 2 (n = 332)*	Stage 3 (n = 195)*
Age	Range	33.4 – 75.6	34 - 70	35-65
	Median	52.7	54	49
	Mean (SD[±])	52.9 (9.4)	52.5 (7.1)	49.4 (7.3)
Height	Range	136.8 – 185.0	139.5 – 177.5	137.5 -176.0
	Median	160	159.3	159
	Mean (SD)	160.1 (6.7)	159.0 (6.4)	158.7 (6.7)
Weight	Range	39.6 – 110.2	42.8 – 106	44 – 103.5
	Median	64.5	63.5	62
	Mean (SD)	66.2 (11.2)	64.4 (11.2)	63.8 (9.8)
BMI[±]				
- **<25 kg/m^2**	n (%)	182 (48.4)	164 (49.4)	98 (50.3)
- **25-30 kg/m^2**	n (%)	141 (37.5)	118 (35.5)	70 (35.9)
- **≥30 kg/m^2**	n (%)	53 (14.1)	50 (15.1)	25 (12.8)
Physical Activity	n (%)	**Inactive:** 34 (9.0)		
	n (%)	**Moderately Inctive:** 85 (22.6)		
	n (%)	**Moderately Active:** 210 (55.9)		
	n (%)	**Active:** 45 (12.0)		
	n (%)	**Missing:** 2 (0.5)		
	Range		1 – 5	0.5 – 30
	Median		3	8.5
	Mean (SD)		2.6 (0.8)	10.0 (6.9)
Smoking Status				
- **Current Smoker**	n (%)	79 (21.1)	69 (20.9)	35 (17.9)
- **Former Smoker**	n (%)	65 (17.4)	66 (20.0)	48 (24.6)
- **Never**	n (%)	230 (61.5)	195 (59.1)	112 (57.5)
Daily alcohol consumption (g/day)	Range	0 – 51.2	0 – 88.7	0 – 62.6
	Median	3.5	1.9	3.3
	Mean (SD)	6.5 (8.4)	8.7 (13.1)	9.6 (13.0)
Daily folate consumption	Range	90.4 – 1113.0	45.3 – 586.2	52.6 – 644.8

(µg/day)				
	Median	268.5	236.1	264.1
	Mean (SD)	291.5 (107.9)	247.3 (82.0)	276.9 (95.4)
Age at Menarche				
- **≤ 11 yrs**	n (%)	72 (19.4)	62 (18.8)	47 (24.1)
- **12 - 14 yrs**	n (%)	242 (65.2)	234 (70.9)	134 (68.7)
- **≥ 15 yrs**	n (%)	57 (15.4)	34 (10.3)	14 (7.2)
Age at FFTP[±]				
- **<25 yrs**	n (%)	147 (45.6)	118 (35.5)	65 (41.9)
- **25 – 30 yrs**	n (%)	131 (40.7)	118 (35.5)	69 (44.5)
- **>30 yrs**	n (%)	44 (13.7)	96 (29.0)	21 (13.6)
Parous				
- **No**	n (%)	45 (12.0)	29 (8.8)	40 (20.5)
- **Yes**	n (%)	330 (88.0)	301 (91.2)	155 (79.5)
Breastfeeding				
- **No**	n (%)	109 (30.0)	108 (32.7)	75 (48.4)
- **Yes**	n (%)	255 (70.0)	222 (67.3)	80 (51.6)
Menopausal Status				
- **Premenopausal**	n (%)	175 (46.5)	155 (46.7)	90 (46.2)
- **Postmenopausal**	n (%)	201 (53.5)	177 (53.3)	105 (53.8)
HRT[±] **use**				
- **Ever**	n (%)	29 (7.8)	13 (3.9)	34 (17.4)
- **Never**	n (%)	343 (92.2)	317 (96.1)	158 (81.0)
OC[±] **use**				
- **Ever**	n (%)	176 (46.9)	131 (39.7)	94 (48.2)
- **Never**	n (%)	199 (53.1)	199 (60.3)	101 (51.8)

*Failure of category counts to add up to this value denotes missing values

± SD: Standard Deviation, BMI: Body Mass Index, FFTP: First Full Term Pregnancy, HRT: Hormone Replacement Therapy, OC: Oral Contraceptive

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

Variable		Methylation Quartile: Cut-offs				
	Units	1: 23.0-68.4 % (n=103)	2: 68.5-71.7 % (n=91)	3: 71.8-74.0 % (n=93)	4: 74.1-80.0 % (n=89)	p-value[a]
Age at blood collection	Mean ± SD	52.9 ± 8.5	52.2 ± 8.5	53.1 ± 10.0	53.6 ± 10.7	0.863
Height	Mean ± SD	159.0 ± 6.1	160.7 ± 6.4	160.8 ± 6.7	160.0 ± 7.5	0.424
Weight	Mean ± SD	66.0 ± 12.3	66.5 ± 10.5	64.8 ± 11.3	67.7 ± 10.4	0.316
BMI [±]	Mean ± SD	26.2 ± 5.1	25.8 ± 4.1	25.1 ± 4.4	26.5 ± 4.4	0.112
Physical Activity						
- **Inactive**	N, (%)	4 (3.9)	9 (9.9)	11 (11.8)	10 (11.2)	
- **Moderately Inactive**	N, (%)	29 (28.2)	22 (24.2)	16 (17.2)	18 (20.3)	
- **Moderately Active**	N, (%)	60 (58.3)	45 (49.4)	54 (58.1)	51 (57.3)	
- **Active**	N, (%)	9 (8.7)	15 (16.5)	11 (11.8)	10 (11.2)	
- **Missing**	N, (%)	1 (0.9)	0 (0.0)	1 (1.1)	0 (0.0)	0.404
Smoking Status						
- **Current Smoker**	N, (%)	58 (56.3)	59 (64.8)	49 (52.7)	47 (52.9)	
- **Former Smoker**	N, (%)	23 (22.3)	12 (13.2)	20 (21.5)	22 (24.7)	
- **Never**	N, (%)	22 (21.4)	20 (22.0)	23 (24.7)	19 (21.3)	
- **Unknown**	N, (%)	0 (0.0)	0 (0.0)	1 (1.1)	1 (1.1)	0.596
Alcohol consumption – lifetime average (g/day)	Mean ± SD	7.0 ± 8.6	6.9 ± 9.7	5.4 ±6.4	6.7 ±8.6	0.822
Dietary folate intake (g/day)	Mean ± SD	294.6 ± 121.9	276.9 ± 96.7	293.1 ± 104.2	301.0 ± 105.3	0.386
Age at menarche	Mean ± SD	13.2 ± 1.7	13.3 ± 1.7	12.5 ± 1.4	12.8 ± 1.7	**0.002***
Age at FFTP[±]	Mean ± SD	25.3 ± 3.5	26.2 ± 3.9	24.8 ± 4.0	25.0 ± 3.7	0.132
Menopausal Status						
- **Pre**	N, (%)	42 (40.8)	47 (51.6)	46 (49.5)	40 (44.9)	
- **Post**	N, (%)	60 (58.3)	43 (47.3)	45 (48.4)	49 (55.1)	
- **Surgical Post**	N, (%)	1 (0.9)	1 (1.1)	2 (2.1)	0 (0.0)	0.545
Parous						
- **No**	N, (%)	12 (11.7)	7 (7.7)	13 (14.0)	19 (21.3)	
- **Yes**	N, (%)	90 (87.4)	84 (92.3)	79 (84.9)	70 (78.7)	
- **Unknown**	N, (%)	1 (0.9)	0 (0.0)	1 (1.1)	0 (0.0)	0.075
Breastfeeding						

- **No**	N, (%)	30 (29.1)	19 (20.9)	26 (28.0)	34 (38.2)	
- **Yes**	N, (%)	70 (68.0)	70 (76.9)	62 (66.7)	53 (59.6)	
- **Unknown**	N, (%)	3 (2.9)	2 (2.2)	5 (5.3)	2 (2.2)	0.086
OC[±] use						
- **No**	N, (%)	50 (48.5)	53 (58.2)	47 (50.5)	49 (55.1)	
- **Yes**	N, (%)	53 (51.5)	38 (41.8)	46 (49.5)	39 (43.8)	
- **Unknown**	N, (%)	0 (0.0)	0 (0.0)	0 (0.0)	1 (1.1)	0.512
HRT[±] use						
- **No**	N, (%)	93 (90.3)	84 (92.3)	81 (87.1)	85 (95.5)	
- **Yes**	N, (%)	3 (2.9)	6 (6.6)	10 (10.7)	4 (4.5)	
- **Unknown**	N, (%)	7 (6.8)	1 (1.1)	2 (2.2)	0 (0.0)	0.399
Highest Education						
- **None**	N, (%)	10 (9.6)	6 (6.6)	7 (7.4)	6 (6.8)	
- **Primary**	N, (%)	45 (43.7)	41 (45.1)	36 (38.7)	35 (39.3)	
- **Technical/ Professional**	N, (%)	15 (14.6)	9 (9.9)	18 (19.4)	20 (22.5)	
- **Secondary**	N, (%)	12 (11.7)	21 (23.1)	17 (18.3)	17 (19.1)	
- **University**	N, (%)	12 (11.7)	13 (14.3)	14 (15.1)	10 (11.2)	
- **Unspecified**	N, (%)	9 (8.7)	1 (1.0)	1 (1.1)	1 (1.1)	0.264

[a] For continuous variables, P-value was derived from Kruskal-Wallis test. For categorical variables, P-value was derived from a chi square test, with the exclusion of “Unknown” categories due to their small cell counts. Both reflect the association between quartiles of methylation and the investigated variables.

*Significant at the Bonferroni-corrected significance cut off (P=0.003) for multiple comparisons

[±] BMI: Body Mass Index, FFTP: First Full Term Pregnancy, HRT: Hormone Replacement Therapy, OC: Oral Contraceptive

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

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

[a] Each OR is adjusted for all other variables in the table

*Significant at the 0.05 level

Table 4 – Significant CpG sites in a linear regression model

TargetID	P-value*	Q-value**	Regression Coefficient***	Chromosome number	Gene	CpG Position Relative to Gene	CpG Island's Name
cg01339004	8.83E-08	0.0392	-0.2765	15	SMAD6	Body	NA

Methylation treated as a continuous outcome (M-values: PBC and COMBAT on chip) and menarcheal age category (>11 vs. ≤11 years) treated as a categorical exposure. Adjusting for age at blood collection, case-control status, and position on the chip.
Analysis of all subjects or of only the 240 subjects that remained healthy for at least 5 years following recruitment yielded the same results.

*P value from a liner regression model where methylation is treated as a continuous outcome (M-values: PBC and COMBAT on chip) and the effect of age at menarche as a categorical variable (>11 vs. ≤11 years), adjusted for age, case-control status, and chip position.
**Q value: False Discovery Rate (FDR) corrected P-value
*** The regression coefficient for each probe; change in methylation for having an age at menarche >11years vs. ≤11 years.

APPENDIX G

Summary of Permission for Third Party Copyright Works

Page Number	Type of work: text, figure, map, etc.	Source work	Copyright holder & year	Work out of copyright	Permission to re-use	Permission requested	permission refused	Orphan work
Page #: Figure 1	Figures 22 and 31	Health, United States, 2011: With Special Feature on SES and Health	U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics ©2011		√			
Page #: Figure 2	Figure 6. Unemployment rates by gender, EU, seasonally adjusted, January 2000 - March 2013 (%)	http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Unemployment_statistics	**© European Union**		√			
Page: # Figure 3	Figure 2. Association of obesity prevalence and age among women of the Diogenes cohort.	Trend in Obesity Prevalence in European Adult Cohort Populations during Follow-up since 1996 and Their Predictions to 2015. *Plos One* 2011; **6**: e27455.	The Authors ©2011		√			
Page: # Figure 4	**From additional file 1, Figure 1: Prevalence of female overweight and obesity, adults ≥ 20 years, by country and year, 1980–2008**	National, regional, and global trends in adult overweight and obesity prevalences. *Popul Heal Metrics* 2012; **10**: 22	The Authors ©2012		√			
Page: # Figure 5	Figure 10. Obesity among US children by age, 1988-1994 through to 2009-2010.	Health, United States, 2011: With Special Feature on SES and Health	U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics ©2011		√			
Page: # Figure 6	Figure 2. Changes in prevalence of overweight and obesity per year.	International Obesity Task Force (IOTF). EU Platform on Diet, Physical Activity and Health: International Obesity Task Force EU platform Briefing Paper. Brussels, IOTF, 2005.	IOTF ©2004		√			
Page: # Figure 7	Figure 1. Mean age at menarche of women born between 1915 and 1964, adjusted for socioeconomic status, for 5-year birth cohorts for each participating country	Age at menarche in relation to adult height: the EPIC study. *Am J Epidemiol* 2005; **162**: 623–32	Copyright © 2005, Oxford University Press		√			

Page: # Figure 8	Figure 2. Average height (cm) at time for conscription per country from 1880 to 2006.	Secular changes in human biological variables in Western countries: an updated review and synthesis. *J Anthr Sci Riv Antropol Jass Ist Ital Antropol* 2008; **86**: 91–112.	Journal of Antrhopological Sciences Copyright © 2008, The Authors		√			
Page: # Figure 9	Figure 4. Mortality among children younger than five years of age and percentage of deprived households (lacking three or more essential items) in selected countries in the WHO European Region.	WHO European review of social determinants of health and the health divide. *Lancet* 2012; 380: 1011–29 Data from UNICEF and Eurostat database. Figure reproduced with permission of Bradshaw J and Mayhew E (University of York, York, UK, personal communication).	Copyright: Bradshaw J and Mayhew E, University of York		√			
Page: # Figure 10	Figure 6. Model of the pathways through which SES influences health	Socioeconomic status and health: what we know and what we don't. *Ann N Y Acad Sci* 1999; **896**: 3–15	Copyright © 2006, John Wiley and Sons		√			
Page: # Figure 11	Figure 1. Folate in one-carbon metabolism. MTHFR sits as a fulcrum between DNA methylation and DNA synthesis, balancing the two pathways to maintain normal homeostasis	Folate status: effects on pathways of colorectal carcinogenesis. *J Nutr* 2002; **132**: 2413S–2418S	Copyright © 2002, American Society for Nutrition		√			
Page: # Figure 12	Figure 1. The predominant effect of ethanol on one-carbon metabolism appears to be the inhibition of methionine synthase.	Effects of alcohol on folate metabolism: implications for carcinogenesis. *Alcohol Fayettev N* 2005; **35**: 235–41.	Copyright © 2005, Elsevier		√			
Page: # Figure 15	Figure 1. Luminometric output: HpaII and MspI digestion with EcoRI followed by LUMA analysis in a Pyrosequencer.	Using LUMA: a Luminometric-based assay for global DNA-methylation. *Epigenetics Off J Dna Methylation Soc* 2006; **1**: 45–8.	©2006 Landes Bioscience		√			
Page: # Figure 16	Infinium Methylation Assay scheme.	High density DNA methylation array with single CpG site resolution. *Genomics* 2011; **98**: 288–95	Copyright © 2011, Elsevier		√			
Page: # Figure 17	Figure 1. Peak-based correction method for improving Infinium II (type II) data which are less accurate and reproducible than Infinium I (type I) data.	Evaluation of the Infinium Methylation 450K technology. *Epigenomics* 2011; **3**: 771–84	Copyright © 2011 François Fuks		√			
Page: #	Figure 4. Two step Mendelian	Two-step epigenetic Mendelian	Copyright © 2012, Oxford		√			

Figure 41	randomization	randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. *Int J Epidemiol* 2012; **41**: 161–76.	University Press					



U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Centers for Disease Control and Prevention
National Center for Health Statistics

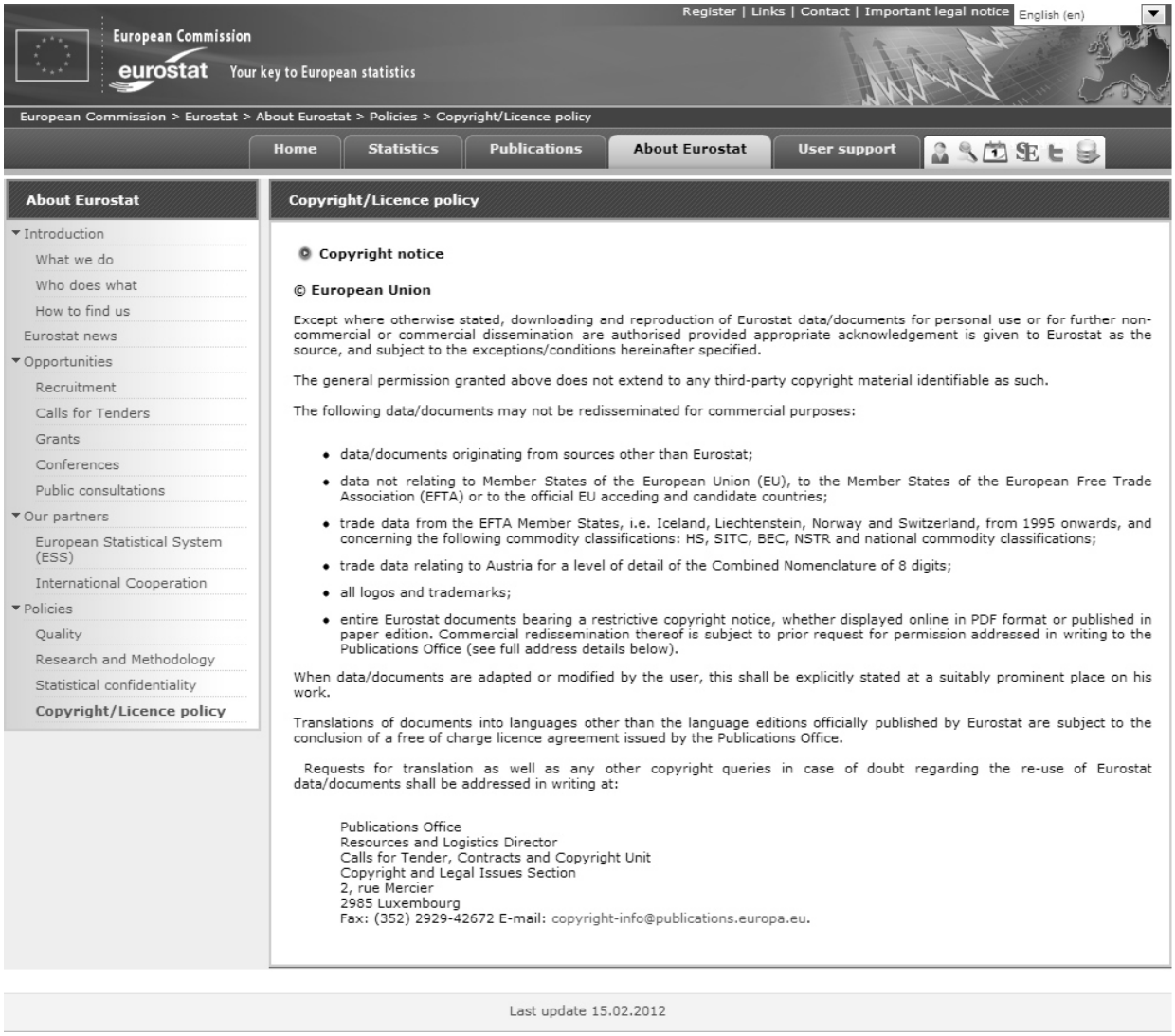
Copyright information

Permission has been obtained from the copyright holders to reproduce certain quoted material in this report. Further reproduction of this material is prohibited without specific permission of the copyright holder. All other material contained in this report is in the public domain and may be used and reprinted without special permission; citation as to source, however, is appreciated.

Suggested citation

National Center for Health Statistics.
Health, United States, 2011: With Special Feature on Socioeconomic Status and Health. Hyattsville, MD. 2012.

Library of Congress Catalog Number 76–641496
For sale by Superintendent of Documents
U.S. Government Printing Office
Washington, DC 20402


Register | Links | Contact | Important legal notice
English (en)
European Commission
eurostat
Your key to European statistics
European Commission > Eurostat > About Eurostat > Policies > Copyright/Licence policy
Home
Statistics
Publications
About Eurostat
User support
About Eurostat
Introduction
What we do
Who does what
How to find us
Eurostat news
Opportunities
Recruitment
Calls for Tenders
Grants
Conferences
Public consultations
Our partners
European Statistical System (ESS)
International Cooperation
Policies
Quality
Research and Methodology
Statistical confidentiality
Copyright/Licence policy
Copyright/Licence policy
Copyright notice
© European Union
Except where otherwise stated, downloading and reproduction of Eurostat data/documents for personal use or for further non-commercial or commercial dissemination are authorised provided appropriate acknowledgement is given to Eurostat as the source, and subject to the exceptions/conditions hereinafter specified.
The general permission granted above does not extend to any third-party copyright material identifiable as such.
The following data/documents may not be redisseminated for commercial purposes:
data/documents originating from sources other than Eurostat;
data not relating to Member States of the European Union (EU), to the Member States of the European Free Trade Association (EFTA) or to the official EU acceding and candidate countries;
trade data from the EFTA Member States, i.e. Iceland, Liechtenstein, Norway and Switzerland, from 1995 onwards, and concerning the following commodity classifications: HS, SITC, BEC, NSTR and national commodity classifications;
trade data relating to Austria for a level of detail of the Combined Nomenclature of 8 digits;
all logos and trademarks;
entire Eurostat documents bearing a restrictive copyright notice, whether displayed online in PDF format or published in paper edition. Commercial rediscemination thereof is subject to prior request for permission addressed in writing to the Publications Office (see full address details below).
When data/documents are adapted or modified by the user, this shall be explicitly stated at a suitably prominent place on his work.
Translations of documents into languages other than the language editions officially published by Eurostat are subject to the conclusion of a free of charge licence agreement issued by the Publications Office.
Requests for translation as well as any other copyright queries in case of doubt regarding the re-use of Eurostat data/documents shall be addressed in writing at:
Publications Office
Resources and Logistics Director
Calls for Tender, Contracts and Copyright Unit
Copyright and Legal Issues Section
2, rue Mercier
2985 Luxembourg
Fax: (352) 2929-42672 E-mail: copyright-info@publications.europa.eu.
Last update 15.02.2012

OPEN ACCESS Freely available online PLoS one

Trend in Obesity Prevalence in European Adult Cohort Populations during Follow-up since 1996 and Their Predictions to 2015

Anne von Ruesten[1]*, Annika Steffen[1], Anna Floegel[1], Daphne L. van der A[2], Giovanna Masala[3], Anne Tjønneland[5], Jytte Halkjaer[5], Domenico Palli[3], Nicholas J. Wareham[4], Ruth J. F. Loos[4], Thorkild I. A. Sørensen[6], Heiner Boeing[1]

1 Department of Epidemiology, German Institute of Human Nutrition Potsdam- Rehbruecke, Nuthetal, Germany, **2** Centre for Nutrition and Health (CVG), National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands, **3** Molecular and Nutritional Epidemiology Unit, Cancer Research and Prevention Institute – ISPO, Florence, Italy, **4** MRC Epidemiology Unit, Institute of Metabolic Science, Cambridge, United Kingdom, **5** Danish Cancer Society, Institute of Cancer Epidemiology, Copenhagen, Denmark, **6** Institute of Preventive Medicine, Copenhagen University Hospital, Copenhagen, Denmark

Abstract

Objective: To investigate trends in obesity prevalence in recent years and to predict the obesity prevalence in 2015 in European populations.

Methods: Data of 97 942 participants from seven cohorts involved in the European Prospective Investigation into Cancer and Nutrition (EPIC) study participating in the Diogenes project (named as "Diogenes cohort" in the following) with weight measurements at baseline and follow-up were used to predict future obesity prevalence with logistic linear and non-linear (leveling off) regression models. In addition, linear and leveling off models were fitted to the EPIC-Potsdam dataset with five weight measures during the observation period to find out which of these two models might provide the more realistic prediction.

Results: During a mean follow-up period of 6 years, the obesity prevalence in the Diogenes cohort increased from 13% to 17%. The linear prediction model predicted an overall obesity prevalence of about 30% in 2015, whereas the leveling off model predicted a prevalence of about 20%. In the EPIC-Potsdam cohort, the shape of obesity trend favors a leveling off model among men ($R^2 = 0.98$), and a linear model among women ($R^2 = 0.99$).

Conclusion: Our data show an increase in obesity prevalence since the 1990ies, and predictions by 2015 suggests a sizeable further increase in European populations. However, the estimates from the leveling off model were considerably lower.

Citation: von Ruesten A, Steffen A, Floegel A, van der A DL, Masala G, et al. (2011) Trend in Obesity Prevalence in European Adult Cohort Populations during Follow-up since 1996 and Their Predictions to 2015. PLoS ONE 6(11): e27455. doi:10.1371/journal.pone.0027455

Editor: Antje Timmer, Bremen Institute of Preventive Research and Social Medicine, Germany

Received July 6, 2011; **Accepted** October 17, 2011; **Published** November 10, 2011

Copyright: © 2011 von Ruesten et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Diogenes is the acronym of the project "Diet, Obesity and Genes" supported by the European Union (Contract no. FP6-513946), https://www.diogenes-eu.org. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

RESEARCH Open Access

National, regional, and global trends in adult overweight and obesity prevalences

Gretchen A Stevens[1], Gitanjali M Singh[2], Yuan Lu[2], Goodarz Danaei[3], John K Lin[2], Mariel M Finucane[2], Adil N Bahalim[4], Russell K McIntire[4], Hialy R Gutierrez[4], Melanie Cowan[5], Christopher J Paciorek[6], Farshad Farzadfar[7], Leanne Riley[5], Majid Ezzati[8*] for the Global Burden of Metabolic Risk Factors of Chronic Diseases Collaborating Group (Body Mass Index)

Abstract

Background: Overweight and obesity prevalence are commonly used for public and policy communication of the extent of the obesity epidemic, yet comparable estimates of trends in overweight and obesity prevalence by country are not available.

Methods: We estimated trends between 1980 and 2008 in overweight and obesity prevalence and their uncertainty for adults 20 years of age and older in 199 countries and territories. Data were from a previous study, which used a Bayesian hierarchical model to estimate mean body mass index (BMI) based on published and unpublished health examination surveys and epidemiologic studies. Here, we used the estimated mean BMIs in a regression model to predict overweight and obesity prevalence by age, country, year, and sex. The uncertainty of the estimates included both those of the Bayesian hierarchical model and the uncertainty due to cross-walking from mean BMI to overweight and obesity prevalence.

Results: The global age-standardized prevalence of obesity nearly doubled from 6.4% (95% uncertainty interval 5.7-7.2%) in 1980 to 12.0% (11.5-12.5%) in 2008. Half of this rise occurred in the 20 years between 1980 and 2000, and half occurred in the 8 years between 2000 and 2008. The age-standardized prevalence of overweight increased from 24.6% (22.7-26.7%) to 34.4% (33.2-35.5%) during the same 28-year period. In 2008, female obesity prevalence ranged from 1.4% (0.7-2.2%) in Bangladesh and 1.5% (0.9-2.4%) in Madagascar to 70.4% (61.9-78.9%) in Tonga and 74.8% (66.7-82.1%) in Nauru. Male obesity was below 1% in Bangladesh, Democratic Republic of the Congo, and Ethiopia, and was highest in Cook Islands (60.1%, 52.6-67.6%) and Nauru (67.9%, 60.5-75.0%).

Conclusions: Globally, the prevalence of overweight and obesity has increased since 1980, and the increase has accelerated. Although obesity increased in most countries, levels and trends varied substantially. These data on trends in overweight and obesity may be used to set targets for obesity prevalence as requested at the United Nations high-level meeting on Prevention and Control of NCDs.

Keywords: Overweight, Obesity, Prevalence, Population health, Risk transition, Global health, Noncommunicable diseases

Find Someone Options

FW: Copyright Permission

EASO Administrator [admin@easo.org]

To: Demetriou, Christiana

05 August 2013 12:

Dear Christiana,

You can use the data. Please note you will need to reference it with 'EASO Childhood Obesity Task Force'.

Kind regards,
Rosie

From: Demetriou, Christiana [mailto:christiana.demetriou08@imperial.ac.uk]
Sent: 29 July 2013 12:01
To: admin@easo.org
Subject: Copyright Permission

Dear Sir/Madam,

I am hoping to include Figure 2 of the attached publication in my PhD thesis which investigates secular trends in obesity and epigenetic health effects.

The figure, with a more detailed legend, is also available under the URL:

http://www.easo.org/childhood-obesity-cotf/facts-a-statistics

What is the procedure for obtaining copyright permission to include this figure in my thesis?

I contacted the IASO but they said that I should contact you instead.

Thank you in advance for your help,

Christiana Demetriou

OXFORD UNIVERSITY PRESS LICENSE
TERMS AND CONDITIONS
May 27, 2013

This is a License Agreement between Christiana A. Demetriou ("You") and Oxford University Press ("Oxford University Press") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Oxford University Press, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	3157100373112
License date	May 27, 2013
Licensed content publisher	Oxford University Press
Licensed content publication	American Journal of Epidemiology
Licensed content title	Age at Menarche in Relation to Adult Height: The EPIC Study
Licensed content author	N. C. Onland-Moret, P. H. M. Peeters, C. H. van Gils, F. Clavel-Chapelon, T. Key, A. Tjønneland, A. Trichopoulou, R. Kaaks, J. Manjer, S. Panico, D. Palli, B. Tehard, M. Stoikidou, H. B. Bueno-De-Mesquita, H. Boeing, K. Overvad, P. Lenner, J. R. Quirós, M. D. Chirlaque, A. B. Miller, K. T. Khaw, E. Riboli
Licensed content date	1 October 2005
Type of Use	Thesis/Dissertation
Institution name	
Title of your work	The impact of secular trends, dietary patterns and socioeconomic conditions on DNA methylation: an epigenome wide investigation
Publisher of your work	n/a
Expected publication date	Oct 2013
Permissions cost	0.00 EUR
Value added tax	0.00 EUR
Total	0.00 EUR
Total	0.00 EUR

Terms and Conditions

STANDARD TERMS AND CONDITIONS FOR REPRODUCTION OF MATERIAL FROM AN OXFORD UNIVERSITY PRESS JOURNAL

1. Use of the material is restricted to the type of use specified in your order details.

2. This permission covers the use of the material in the English language in the following territory: world. If you have requested additional permission to translate this material, the terms and conditions of this reuse will be set out in clause 12.

3. This permission is limited to the particular use authorized in (1) above and does not allow you to sanction its use elsewhere in any other format other than specified above, nor does it apply to quotations, images, artistic works etc that have been reproduced from other sources which may be part of the material to be used.

4. No alteration, omission or addition is made to the material without our written consent. Permission must be re-cleared with Oxford University Press if/when you decide to reprint.

5. The following credit line appears wherever the material is used: author, title, journal, year, volume, issue number, pagination, by permission of Oxford University Press or the sponsoring society if the journal is a society journal. Where a journal is being published on behalf of a learned society, the details of that society must be included in the credit line.

6. For the reproduction of a full article from an Oxford University Press journal for whatever purpose, the corresponding author of the material concerned should be informed of the proposed use. Contact details for the corresponding authors of all Oxford University Press journal contact can be found alongside either the abstract or full text of the article concerned, accessible from www.oxfordjournals.org Should there be a problem clearing these rights, please contact journals.permissions@oxfordjournals.org

7. If the credit line or acknowledgement in our publication indicates that any of the figures, images or photos was reproduced, drawn or modified from an earlier source it will be necessary for you to clear this permission with the original publisher as well. If this permission has not been obtained, please note that this material cannot be included in your publication/photocopies.

8. While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the

transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by Oxford University Press or by Copyright Clearance Center (CCC)) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and Oxford University Press reserves the right to take any and all action to protect its copyright in the materials.

9. This license is personal to you and may not be sublicensed, assigned or transferred by you to any other person without Oxford University Press's written permission.

10. Oxford University Press reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

11. You hereby indemnify and agree to hold harmless Oxford University Press and CCC, and their respective officers, directors, employs and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

12. Other Terms and Conditions:

v1.4

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number RLNK501029831.
Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.

Make Payment To:
Copyright Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006

For suggestions or comments regarding this order, contact RightsLink Customer Support: customercare@copyright.com or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.

Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.

isita@isita-org.com

29 May 2013 09:57

Hi Christiana, please write to Marica Danubio (in cc), as the Editor of the JASs I am happy that someone uses our material.

I will be happy to get through your work when you feel it appropriate, the topic sounds very interesting.

thanks for your correctness.

regards

Demetriou, Christiana

Sent Items

29 May 2013 13:24

Dear Marica Danubio,

I am hoping to reuse Figure 2 of the attached publication in my PhD thesis, investigating secular trends and their impact on DNA methylation.

I have e-mailed the editor of the journal and received her permission (see e-mail below).

Would it be ok also with you for me to use this figure?

Thank you in advance,
Christiana Demetriou

mariaenrica.danubio@cc.univaq.it on behalf of Marica Danubio [marica.dan... Actions

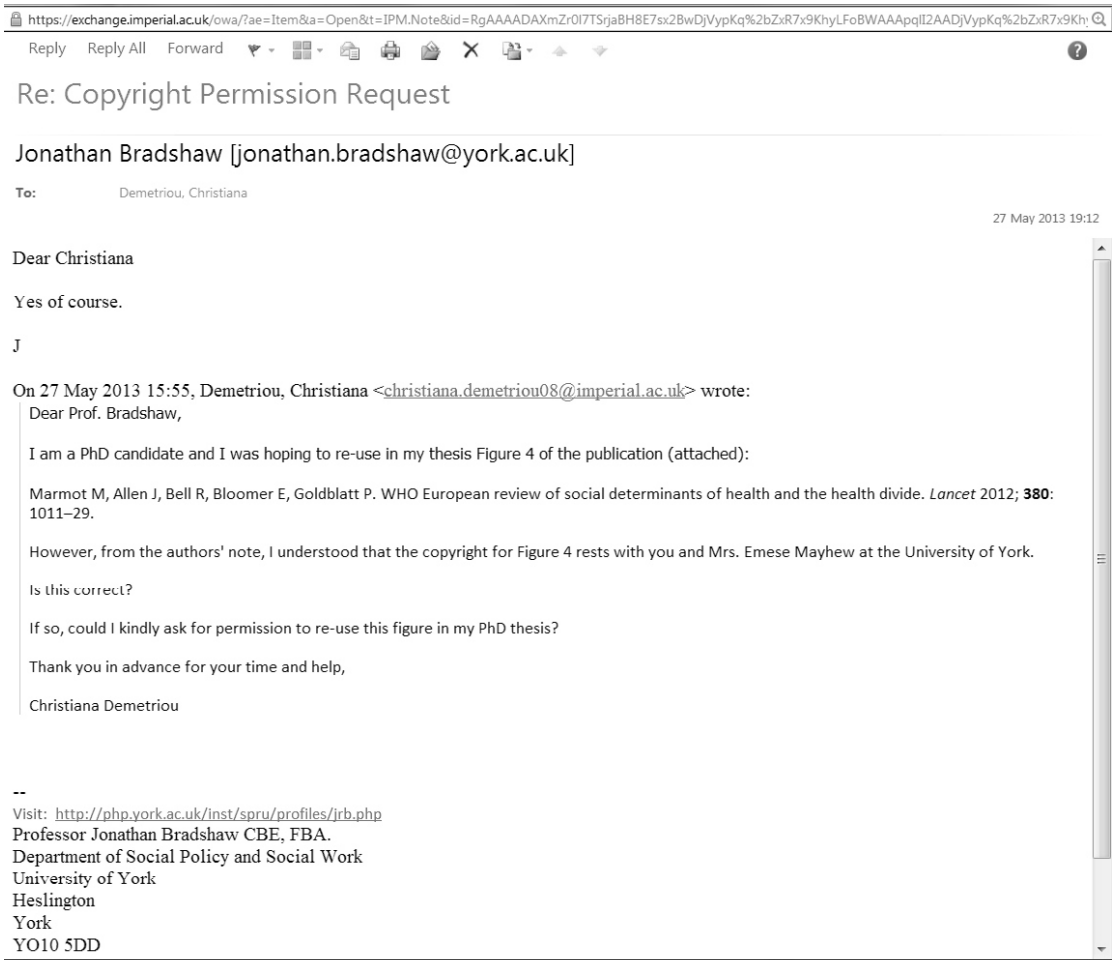
In response to the message from Demetriou, Christiana, 29/05/2013

To: Demetriou, Christiana

30 May 2013 16:08

- You replied on 30/05/2013 16:19.

Dear Christiana Demetriou,
yes, of course you can use Figure 2.
Best wishes,
Marica Danubio



https://exchange.imperial.ac.uk/owa/?ae=Item&a=Open&t=IPM.Note&id=RgAAAADAXmZr0l7TSrjaBH8E7sx2BwDjVypKq%2bZxR7x9KhyLFoBWAAApqII2AADjVypKq%2bZxR7x9Kh

Reply Reply All Forward

Re: Copyright Permission Request

Jonathan Bradshaw [jonathan.bradshaw@york.ac.uk]

To: Demetriou, Christiana

27 May 2013 19:12

Dear Christiana

Yes of course.

J

On 27 May 2013 15:55, Demetriou, Christiana <christiana.demetriou08@imperial.ac.uk> wrote:

> Dear Prof. Bradshaw,
>
> I am a PhD candidate and I was hoping to re-use in my thesis Figure 4 of the publication (attached):
>
> Marmot M, Allen J, Bell R, Bloomer E, Goldblatt P. WHO European review of social determinants of health and the health divide. *Lancet* 2012; **380**: 1011–29.
>
> However, from the authors' note, I understood that the copyright for Figure 4 rests with you and Mrs. Emese Mayhew at the University of York.
>
> Is this correct?
>
> If so, could I kindly ask for permission to re-use this figure in my PhD thesis?
>
> Thank you in advance for your time and help,
>
> Christiana Demetriou

--

Visit: http://php.york.ac.uk/inst/spru/profiles/jrb.php
Professor Jonathan Bradshaw CBE, FBA.
Department of Social Policy and Social Work
University of York
Heslington
York
YO10 5DD

JOHN WILEY AND SONS LICENSE
TERMS AND CONDITIONS
May 27, 2013

This is a License Agreement between Christiana A. Demetriou ("You") and John Wiley and Sons ("John Wiley and Sons") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by John Wiley and Sons, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	3157120280207
License date	May 27, 2013
Licensed content publisher	John Wiley and Sons
Licensed content publication	Annals of the New York Academy of Sciences
Licensed content title	Socioeconomic Status and Health: What We Know and What We Don't
Licensed copyright line	Copyright © 2006, John Wiley and Sons
Licensed content author	NANCY E. ADLER,JOAN M. OSTROVE
Licensed content date	Feb 6, 2006
Start page	3
End page	15
Type of use	Dissertation/Thesis
Requestor type	University/Academic
Format	Print and electronic
Portion	Figure/table
Number of figures/tables	1
Original Wiley figure/table number(s)	Figure 6
Will you be translating?	No
Total	0.00 USD
Terms and Conditions	

TERMS AND CONDITIONS

This copyrighted material is owned by or exclusively licensed to John Wiley & Sons, Inc. or one of its group companies (each a "Wiley Company") or a society for whom a Wiley Company has exclusive publishing rights in relation to a particular journal (collectively "WILEY"). By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the billing and payment terms and conditions established by the Copyright Clearance Center Inc., ("CCC's Billing and Payment terms and conditions"), at the time that you opened your RightsLink account (these are available at any time at http://myaccount.copyright.com).

Terms and Conditions

1. The materials you have requested permission to reproduce (the "Materials") are protected by copyright.
2.You are hereby granted a personal, non-exclusive, non-sublicensable, non-transferable, worldwide, limited license to reproduce the Materials for the purpose specified in the licensing process. This license is for a one-time use only with a maximum distribution equal to the number that you identified in the licensing process. Any form of republication granted by this license must be completed within two years of the date of the grant of this license (although copies prepared before may be distributed thereafter). The Materials shall not be used in any other manner or for any other purpose. Permission is granted subject to an appropriate acknowledgement given to the author, title of the material/book/journal and the publisher. You shall also duplicate the copyright notice that appears in the Wiley publication in your use of the Material. Permission is also granted on the understanding that nowhere in the text is a previously published source acknowledged for all or part of this Material. Any third party material is expressly excluded from this permission.
3. With respect to the Materials, all rights are reserved. Except as expressly granted by the terms of the license, no part of the Materials may be copied, modified, adapted (except for minor reformatting required by the new Publication), translated, reproduced, transferred or distributed, in any form or by any means, and no derivative works may be made based on the Materials without the prior permission of the respective copyright owner. You may not alter, remove or suppress in any manner any copyright, trademark or other notices displayed by the Materials. You may not license, rent, sell, loan, lease, pledge, offer as security, transfer or assign the Materials, or any of the rights granted to you hereunder to any other person.
4. The Materials and all of the intellectual property rights therein shall at all times remain the exclusive property of John Wiley & Sons Inc or one of its related companies (WILEY) or their respective licensors, and your interest therein is only that of having possession of and the right to reproduce the Materials pursuant to Section 2 herein during the continuance of this Agreement. You agree that you own no right, title or interest in or to the Materials or any of the intellectual property rights therein. You shall have no rights hereunder other than the license as provided for above in Section 2. No right, license or interest to any trademark, trade name, service mark or other branding ("Marks") of WILEY or its licensors is granted hereunder, and you agree that you shall not assert any such right, license or interest with respect thereto.
5. NEITHER WILEY NOR ITS LICENSORS MAKES ANY WARRANTY OR REPRESENTATION OF ANY KIND TO YOU OR ANY THIRD PARTY, EXPRESS, IMPLIED OR STATUTORY, WITH RESPECT TO THE MATERIALS OR THE ACCURACY OF ANY INFORMATION CONTAINED IN THE MATERIALS, INCLUDING, WITHOUT LIMITATION, ANY IMPLIED WARRANTY OF MERCHANTABILITY, ACCURACY, SATISFACTORY QUALITY, FITNESS FOR A PARTICULAR PURPOSE, USABILITY, INTEGRATION OR NON-INFRINGEMENT AND ALL SUCH WARRANTIES ARE HEREBY EXCLUDED BY WILEY AND ITS LICENSORS AND WAIVED BY YOU.
6. WILEY shall have the right to terminate this Agreement immediately upon breach of this Agreement by you.
7. You shall indemnify, defend and hold harmless WILEY, its Licensors and their respective directors, officers, agents and employees, from and against any actual or threatened claims, demands, causes of action or proceedings arising from any breach of this Agreement by you.
8. IN NO EVENT SHALL WILEY OR ITS LICENSORS BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR ENTITY FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL, INDIRECT, EXEMPLARY OR PUNITIVE DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, PROVISIONING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.
9. Should any provision of this Agreement be held by a court of competent jurisdiction to be illegal, invalid, or unenforceable, that provision shall be deemed amended to achieve as nearly as possible the same economic effect as the original provision, and the legality, validity and enforceability of the remaining provisions of this Agreement shall not be affected or impaired thereby.
10. The failure of either party to enforce any term or condition of this Agreement shall not constitute a waiver of either party's right to enforce each and every term and condition of this Agreement. No breach under this agreement shall be deemed waived or excused by either party unless such waiver or consent is in writing signed by the party granting such waiver or consent. The waiver by or consent of a party to a breach of any provision of this Agreement shall not operate or be construed as a waiver of or consent to any other or subsequent breach by such other party.
11. This Agreement may not be assigned (including by operation of law or otherwise) by you without WILEY's prior written consent.
12. Any fee required for this permission shall be non-refundable after thirty (30) days from receipt
13. These terms and conditions together with CCC's Billing and Payment terms and conditions (which are incorporated herein) form the entire agreement between you and WILEY concerning this licensing transaction and (in the absence of fraud) supersedes all prior agreements and representations of the parties, oral or written. This Agreement may not be amended except in writing signed by both parties. This Agreement shall be binding upon and inure to the benefit of the parties' successors, legal representatives, and authorized assigns.
14. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall prevail.
15. WILEY expressly reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.
16. This Agreement will be void if the Type of Use, Format, Circulation, or Requestor Type was misrepresented during the licensing process.
17. This Agreement shall be governed by and construed in accordance with the laws of the State of New York, USA, without regards to such state's conflict of law rules. Any legal action, suit or proceeding arising out of or relating to these Terms and Conditions or the breach thereof shall be instituted in a court of competent jurisdiction in New York County in the State of New York in the United States of America and each party hereby consents and submits to the personal jurisdiction of such court, waives any objection to venue in such court and consents to service of process by registered or certified mail, return receipt requested, at the last known address of such party.

Wiley Open Access Terms and Conditions

Wiley publishes Open Access articles in both its Wiley Open Access Journals program [http://www.wileyopenaccess.com/view/index.html] and as Online Open articles in its subscription journals. The majority of Wiley Open Access Journals have adopted the Creative Commons Attribution License (CC BY) which permits the unrestricted use, distribution, reproduction, adaptation and commercial exploitation of the article in any medium. No permission is required to use the article in this way provided that the article is properly cited and other license terms are observed. A small number of Wiley Open Access journals have retained the Creative Commons Attribution Non Commercial License (CC BY-NC), which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

Online Open articles - Authors selecting Online Open are, unless particular exceptions apply, offered a choice of Creative Commons licenses. They may therefore select from the CC BY, the CC BY-NC and the Attribution-NoDerivatives (CC BY-NC-ND). The CC BY-NC-ND is more restrictive than the CC BY-NC as it does not permit adaptations or modifications without rights holder consent.
Wiley Open Access articles are protected by copyright and are posted to repositories and websites in accordance with the terms of the applicable Creative Commons license referenced on the article. At the time of deposit, Wiley Open Access articles include all changes made during peer review, copyediting, and publishing. Repositories and websites that host the article are responsible for incorporating any publisher-supplied amendments or retractions issued subsequently.
Wiley Open Access articles are also available without charge on Wiley's publishing platform, **Wiley Online Library** or any successor sites.
Conditions applicable to all Wiley Open Access articles:

- The authors' moral rights must not be compromised. These rights include the right of "paternity" (also known as "attribution" - the right for the author to be identified as such) and "integrity" (the right for the author not to have the work altered in such a way that the author's reputation or integrity may be damaged).
- Where content in the article is identified as belonging to a third party, it is the obligation of the user to ensure that any reuse complies with the copyright policies of the owner of that content.
- If article content is copied, downloaded or otherwise reused for research and other purposes as permitted, a link to the appropriate bibliographic citation (authors, journal, article title, volume, issue, page numbers, DOI and the link to the definitive published version on Wiley Online Library) should be maintained. Copyright notices and disclaimers must not be deleted.
    - Creative Commons licenses are copyright licenses and do not confer any other rights, including but not limited to trademark or patent rights.
- Any translations, for which a prior translation agreement with Wiley has not been agreed, must prominently display the statement: "This is an unofficial translation of an article that appeared in a Wiley publication. The publisher has not endorsed this translation."
  Conditions applicable to non-commercial licenses (CC BY-NC and CC BY-NC-ND)
  For non-commercial and non-promotional purposes individual non-commercial users may access, download, copy, display and redistribute to colleagues Wiley Open Access articles. In addition, articles adopting the CC BY-NC may be adapted, translated, and text- and data-mined subject to the conditions above.
  Use by commercial "for-profit" organizations
  Use of non-commercial Wiley Open Access articles for commercial, promotional, or marketing purposes requires further explicit permission from Wiley and will be subject to a fee. Commercial purposes include:
    - Copying or downloading of articles, or linking to such articles for further redistribution, sale or licensing;
    - Copying, downloading or posting by a site or service that incorporates advertising with such content;
    - The inclusion or incorporation of article content in other works or services (other than normal quotations with an appropriate citation) that is then available for sale or licensing, for a fee (for example, a compilation produced for marketing purposes, inclusion in a sales pack)
    - Use of article content (other than normal quotations with appropriate citation) by for-profit organizations for promotional purposes
    - Linking to article content in e-mails redistributed for promotional, marketing or educational purposes;
    - Use for the purposes of monetary reward by means of sale, resale, license, loan, transfer or other form of commercial exploitation such as marketing products
    - Print reprints of Wiley Open Access articles can be purchased from: corporatesales@wiley.com
      The modification or adaptation for any purpose of an article referencing the CC BY-NC-ND License requires consent which can be requested from RightsLink@wiley.com .

      Other Terms and Conditions:
      BY CLICKING ON THE "I AGREE..." BOX, YOU ACKNOWLEDGE THAT YOU HAVE READ AND FULLY UNDERSTAND EACH OF THE SECTIONS OF AND PROVISIONS SET FORTH IN THIS AGREEMENT AND THAT YOU ARE IN AGREEMENT WITH AND ARE WILLING TO ACCEPT ALL OF YOUR OBLIGATIONS AS SET FORTH IN THIS AGREEMENT.

      v1.8

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number RLNK501029849.
Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.

Make Payment To:
Copyright Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006

For suggestions or comments regarding this order, contact RightsLink Customer Support: customercare@copyright.com or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.

Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.

AMERICAN SOCIETY FOR NUTRITION LICENSE
TERMS AND CONDITIONS
May 28, 2013

This is a License Agreement between Christiana A. Demetriou ("You") and American Society for Nutrition ("American Society for Nutrition") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by American Society for Nutrition, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	3157490641348
License date	May 28, 2013
Volume number	132
Issue number	8
Licensed content publisher	American Society for Nutrition
Licensed content publication	The Journal of Nutrition
Licensed content title	Folate Status: Effects on Pathways of Colorectal Carcinogenesis
Licensed content author	Sang-Woon Choi, Joel B. Mason
Licensed content date	Aug 1, 2002
Type of Use	Thesis/Dissertation
Requestor type	Student
Portion	Figures/table/illustration
Number of Figures/table/illustration	1
List of figures/table/ illustration	Figure 1 - Folate in one-carbon metabolism. MTHFR sits as a fulcrum between DNA methylation and DNA synthesis, balancing the two pathways to maintain normal homeostasis.
Order reference number	
Title of your dissertation / thesis	The impact of secular trends, dietary patterns and socioeconomic conditions on DNA methylation: an epigenome wide investigation
Expected completion date	Oct 2013
Estimated size(pages)	250
Total	0.00 EUR
Terms and Conditions	

Terms and Conditions for Rightslink Licenses

1) This license agreement ("Agreement") is between the Customer as identified by the user in Rightslink ("Customer") and The American Society for Nutrition, Inc, a nonprofit corporation with offices at 9650 Rockville Pike, Bethesda, Maryland 20814-3998 USA ("ASN"), regarding content published in one of the following ASN journals: *The American Journal of Clinical Nutrition (AJCN)* or *The Journal of Nutrition (JN) or Advances in Nutrition*. The license administrator for this Agreement is the Copyright Clearance Center (CCC).

2) Introduction The publisher of this copyrighted material is ASN. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by CCC, at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com.

3) ASN occasionally publishes material for which it does not hold copyright; in such instances, ASN identifies and properly cites the appropriate copyright holder. It is the Customer's responsibility to verify that ASN holds copyright for which the Customer is requesting permission. If the material is credited to another source, then ASN does not have the authority to grant permission for its reuse and this license is to be considered invalid.

4) Limited License ASN hereby grants to you a non-exclusive license to use this material. Licenses are for one-time use only with a maximum distribution equal to the number that you identified in the licensing process. Except when material is licensed for reuse in a scholarly journal, magazine, newsletter, newspaper, book, or textbook, in which case permission extends to reprints of the product produced for publisher restocking, any form of republication must be completed within one year from the date hereof (although copies prepared before then may be distributed thereafter). Reprints must not be new products nor marketed as such: **reuse in new volumes or new editions must be licensed explicitly**. Any form of republication must be completed within one year from the date hereof (although copies prepared before then may be distributed thereafter); and any electronic posting is limited to the time specified in this request, if any. ASN asks that

commercial organizations limit distribution of photocopies to within the organization requesting permission. Requests to distribute content copyrighted by ASN outside of the requesting organization should be processed as reprints or ePrints. Use of more than 50% of an article copyrighted by ASN in a single new work must be reviewed separately by ASN staff.

5) Geographic Rights Licenses may be exercised solely in the country or countries specified by the Customer on the Quick Price or Additional Data page in Rightslink. However, in cases in which ASN has exclusive republication licensees in specific countries, licenses to reproduce full-text material in these countries may be revoked by ASN. In such cases, the Customer would be referred to the license agent in the respective country or countries. For questions about whether this would apply to your order, please contact ASN at publications@nutrition.org.

6) Alterations/Modifications Except for adaptations of tables and figures and for translations of content as described below, material may not be modified. Translations of figures, tables, and excerpts are permitted as transactions within Rightslink. Translations of full-text materials must be approved by ASN before the Customer can proceed with the Rightslink transaction. Full-text translation requests processed through CCC will be considered invalid unless the Customer enters the ASN-issued approval code for the translation. A copy of all full-text translations must be sent to ASN.

7) Reservation of Rights ASN reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions, and (iii) CCC's Billing and Payment terms and conditions.

8) License Contingent on Payment While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by ASN or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.

9) Copyright Notice: Disclaimer You must give complete credit to the original source in connection with any reproduction or adaptation of the licensed material: 'Am J Clin Nutr (year;volume:page range), American Society for Nutrition,' or 'J Nutr (year;volume:page range), American Society for Nutrition;.'Adv Nutr (year;volume:page range), American Society for Nutrition.' For translations: Figures or tables cannot be used in advertisements.

Full-text translations of articles published in *The Journal of Nutrition* must include the following disclaimer: *Translated from the original into LANGUAGE by TRANSLATOR'S NAME. The translator assumes responsibility for the accuracy of the translation. The American Society for Nutrition is not responsible for translation errors. Readers are encouraged to access the original publication at http://jn.nutrition.org.* For photocopies and in-house reprints:

The article will be reprinted in its entirety without change.

No material will be attached to the reprints, and the reprint will be used for educational purposes only, not to promote, sell, or in any way endorse a product.

The citation appearing on the first page of the article must appear in the reprints.

As a courtesy, please notify the author of your intended reuse of his or her content.

All reprints must have reprinted on each: *The American Society for Nutrition, Inc., does not endorse any commercial enterprise.*

10) Warranties ASN makes no representations or warranties with respect to the licensed material.

11) Indemnity You hereby indemnify and agree to hold harmless ASN and CCC, and their respective officers, directors, employees, and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

12) No Transfer of License This license is personal to you, but may be assigned or transferred by you to a business associate (or to your employer) if you give prompt written notice of the assignment or transfer to ASN. No such assignment or transfer shall relieve you of the obligation to pay the designated license fee on a timely basis (although payment by the identified

assignee can fulfill your obligation).

13) No Amendment Except in Writing This license may not be amended except in a writing signed by both parties (or, in the case of ASN, by CCC on ASN's behalf).

14) Objection to Contrary Terms ASN hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and ASN (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

15) Jurisdiction This license Agreement contains the entire understanding of the parties with respect to the licensed content and can be modified only by a signed, written agreement. This Agreement shall be construed in accordance with the laws of the State of Maryland and the US copyright laws. If any term of this Agreement shall be found invalid by any court of competent jurisdiction, such provision shall be enforced to the fullest extent that it is valid and enforceable under applicable law, and all other provisions of this Agreement shall remain in full force and effect. For state and local governments, the Terms and Conditions do not apply when in conflict with existing statues.

16) Other There are additional terms and conditions, established by Copyright Clearance Center, Inc. ("CCC") as the administrator of this licensing service that relate to billing and payment for licenses provided through this service. Those terms and conditions apply to each transaction as if they were restated here. As a user of this service, you agreed to those terms and conditions at the time that you established your account, and you may see them again at any time at http://myaccount.copyright.com.

Special Terms: <%=specialTerms%>

v 2.5

Other Terms and Conditions

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number RLNK501030148.
Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.

Make Payment To:
Copyright Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006

For suggestions or comments regarding this order, contact RightsLink Customer Support:
customercare@copyright.com or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.

Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.

ELSEVIER LICENSE
TERMS AND CONDITIONS
May 28, 2013

This is a License Agreement between Christiana A. Demetriou ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard,Langford Lane Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	Christiana A. Demetriou
Customer address	57 Westcott Crescent London, W7 1PL
License number	3157491197640
License date	May 28, 2013
Licensed content publisher	Elsevier
Licensed content publication	Alcohol
Licensed content title	Effects of alcohol on folate metabolism: implications for carcinogenesis
Licensed content author	Joel B. Mason,Sang-Woon Choi
Licensed content date	April 2005
Licensed content volume number	35
Licensed content issue number	3
Number of pages	7
Start Page	235
End Page	241
Type of Use	reuse in a thesis/dissertation
Intended publisher of new work	other
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Order reference number	
Title of your thesis/dissertation	The impact of secular trends, dietary patterns and socioeconomic conditions on DNA methylation: an epigenome wide investigation
Expected completion date	Oct 2013
Estimated size (number of pages)	250
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 EUR
VAT/Local Sales Tax	0.0 USD / 0.0 GBP
Total	0.00 EUR
Terms and Conditions	

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.

3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."

4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.

5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com)

6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.

7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.

9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.

10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.

12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).

13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In

no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:

15. **Translation**: This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article. If this license is to re-use 1 or 2 figures then permission is granted for non-exclusive world rights in all languages.

16. **Website**: The following terms and conditions apply to electronic reserve and author websites:
Electronic reserve: If licensed material is to be posted to website, the web site is to be password-protected and made available only to bona fide students registered on a relevant course if:
This license was made in connection with a course,
This permission is granted for 1 year only. You may obtain a license for future website posting,
All content posted to the web site must maintain the copyright information line on the bottom of each image,
A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or the Elsevier homepage for books at http://www.elsevier.com , and
Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

17. **Author website** for journals with the following additional clauses:

All content posted to the web site must maintain the copyright information line on the bottom of each image, and the permission granted is limited to the personal version of your paper. You are not allowed to download and post the published electronic version of your article (whether PDF or HTML, proof or final version), nor may you scan the printed edition to create an electronic version. A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx . As part of our normal production process, you will receive an e-mail notice when your article appears on Elsevier's online service ScienceDirect (www.sciencedirect.com). That e-mail will include the article's Digital Object Identifier (DOI). This number provides the electronic link to the published article and should be included in the posting of your personal version. We ask that you wait until you receive this e-mail and have the DOI to do any posting.

Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

18. **Author website** for books with the following additional clauses:
Authors are permitted to place a brief summary of their work online only.
A hyper-text must be included to the Elsevier homepage at http://www.elsevier.com . All content posted to the web site must maintain the copyright information line on the bottom of each image. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version.

Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

19. **Website** (regular and for author): A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx. or for books to the Elsevier homepage at http://www.elsevier.com

20. **Thesis/Dissertation**: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission.

21. **Other Conditions**:

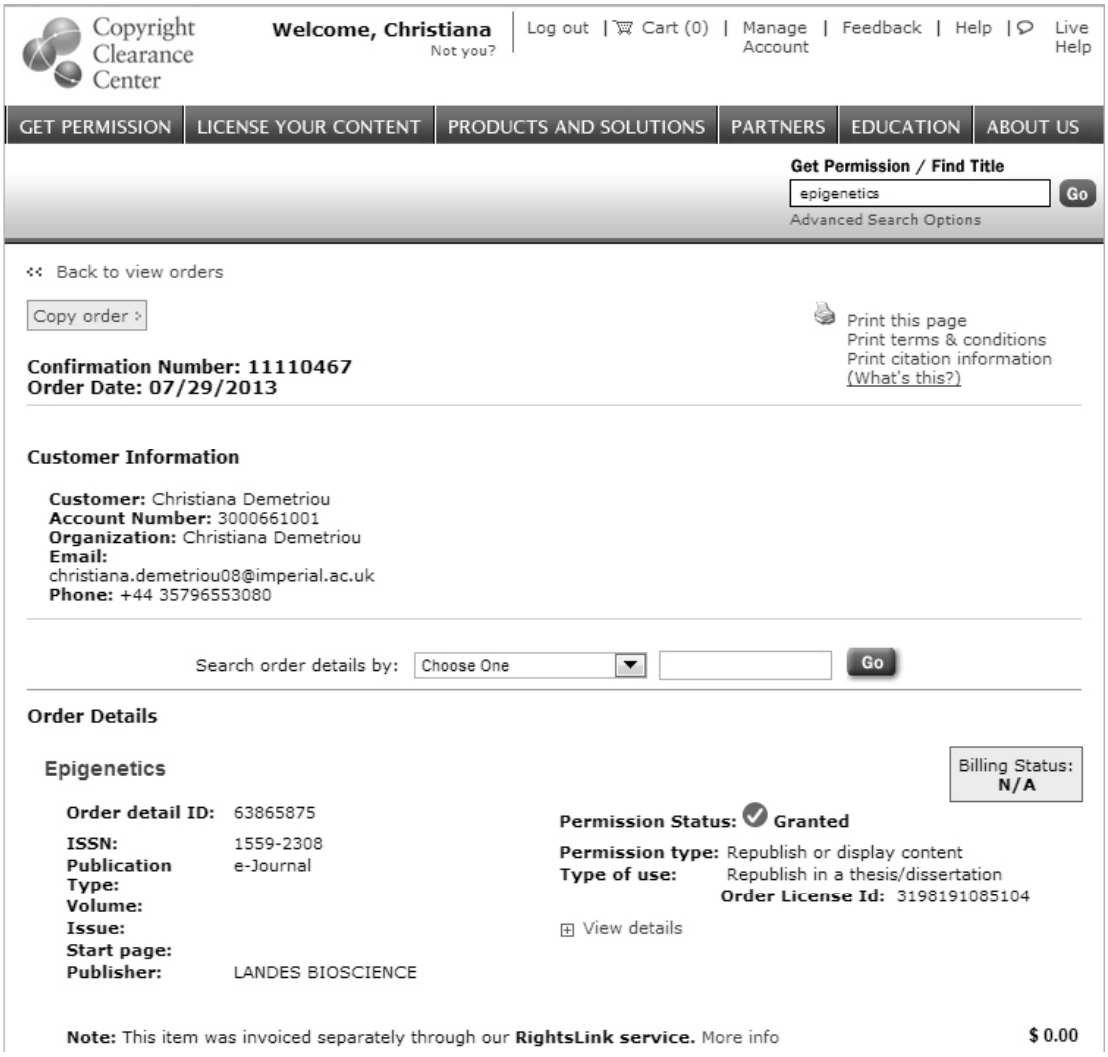
v1.6

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number RLNK501030155.
Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.

Make Payment To:
Copyright Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006

For suggestions or comments regarding this order, contact RightsLink Customer Support:
customercare@copyright.com or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.

Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.


Copyright Clearance Center
Welcome, Christiana
Not you?
Log out | Cart (0) | Manage Account | Feedback | Help | Live Help
GET PERMISSION
LICENSE YOUR CONTENT
PRODUCTS AND SOLUTIONS
PARTNERS
EDUCATION
ABOUT US
Get Permission / Find Title
epigenetics
Go
Advanced Search Options
Back to view orders
Copy order
Print this page
Print terms & conditions
Print citation information
(What's this?)
Confirmation Number: 11110467
Order Date: 07/29/2013
Customer Information
Customer: Christiana Demetriou
Account Number: 3000661001
Organization: Christiana Demetriou
Email:
christiana.demetriou08@imperial.ac.uk
Phone: +44 35796553080
Search order details by:
Choose One
Go
Order Details
Epigenetics
Billing Status:
N/A
Order detail ID: 63865875
ISSN: 1559-2308
Publication Type: e-Journal
Volume:
Issue:
Start page:
Publisher: LANDES BIOSCIENCE
Permission Status: Granted
Permission type: Republish or display content
Type of use: Republish in a thesis/dissertation
Order License Id: 3198191085104
View details
Note: This item was invoiced separately through our RightsLink service. More info
$ 0.00

ELSEVIER LICENSE
TERMS AND CONDITIONS
May 28, 2013

This is a License Agreement between Christiana A. Demetriou ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard,Langford Lane Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	Christiana A. Demetriou
Customer address	57 Westcott Crescent London, W7 1PL
License number	3157500434084
License date	May 28, 2013
Licensed content publisher	Elsevier
Licensed content publication	Genomics
Licensed content title	High density DNA methylation array with single CpG site resolution
Licensed content author	Marina Bibikova,Bret Barnes,Chan Tsan,Vincent Ho,Brandy Klotzle,Jennie M. Le,David Delano,Lu Zhang,Gary P. Schroth,Kevin L. Gunderson,Jian-Bing Fan,Richard Shen
Licensed content date	October 2011
Licensed content volume number	98
Licensed content issue number	4
Number of pages	8
Start Page	288
End Page	295
Type of Use	reuse in a thesis/dissertation
Intended publisher of new work	other
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Order reference number	
Title of your thesis/dissertation	The impact of secular trends, dietary patterns and socioeconomic conditions on DNA methylation: an epigenome wide investigation
Expected completion date	Oct 2013
Estimated size (number of pages)	250
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 EUR
VAT/Local Sales Tax	0.0 USD / 0.0 GBP

Total 0.00 EUR

Terms and Conditions

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.

3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."

4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.

5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com)

6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.

7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.

9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.

10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the

licensed material other than as specifically authorized pursuant to this license.

11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.

12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).

13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:

15. **Translation**: This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article. If this license is to re-use 1 or 2 figures then permission is granted for non-exclusive world rights in all languages.

16. **Website**: The following terms and conditions apply to electronic reserve and author websites:
Electronic reserve: If licensed material is to be posted to website, the web site is to be password-protected and made available only to bona fide students registered on a relevant course if:
This license was made in connection with a course,
This permission is granted for 1 year only. You may obtain a license for future website posting,
All content posted to the web site must maintain the copyright information line on the bottom of each image,
A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or the Elsevier homepage for books at http://www.elsevier.com , and
Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

17. **Author website** for journals with the following additional clauses:

All content posted to the web site must maintain the copyright information line on the bottom of each image, and the permission granted is limited to the personal version of your paper. You are not allowed to download and post the published electronic version of your article (whether PDF or HTML, proof or final version), nor may you scan the printed edition to create an electronic version. A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx . As part of our normal production process, you will receive an e-mail notice when your article appears on Elsevier's online service ScienceDirect (www.sciencedirect.com). That e-mail will include the article's Digital Object Identifier (DOI). This

number provides the electronic link to the published article and should be included in the posting of your personal version. We ask that you wait until you receive this e-mail and have the DOI to do any posting.

Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

18. **Author website** for books with the following additional clauses:
Authors are permitted to place a brief summary of their work online only.
A hyper-text must be included to the Elsevier homepage at http://www.elsevier.com . All content posted to the web site must maintain the copyright information line on the bottom of each image. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version.

Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

19. **Website** (regular and for author): A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx. or for books to the Elsevier homepage at http://www.elsevier.com

20. **Thesis/Dissertation**: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission.

21. **Other Conditions**:

v1.6

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number RLNK501030176.
Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.

Make Payment To:
Copyright Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006

For suggestions or comments regarding this order, contact RightsLink Customer Support: customercare@copyright.com or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.

Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.

Permission to use Future Medicine Ltd copyright material

Request from:

- Contact name: Christiana Demetriou
- Publisher/company name:
- Address:
- Telephone/e-mail: christiana.demetriou08@imperial.ac.uk

Request details:

- Request to use the following content: Figure 1 - Epigenomics - December 2011, Vol. 3, No. 6, Pages 771-784
- In the following publication: Thesis
- In what media (print/electronic/print & electronic): Print and Electronic
- In the following languages: All

We, Future Medicine Ltd, grant permission to reuse the material specified above within the publication specified above.

Notes and conditions:

1. This permission is granted free of charge, for one-time use only.
2. Future Medicine Ltd grant the publisher non-exclusive world rights to publish the content in the publication/website specified above.
3. Future Medicine Ltd retains copyright ownership of the content.
4. Permission is granted on a one-time basis only. Separate permission is required for any further use or edition.
5. The publisher will make due acknowledgement of the original publication wherever they republish the content: citing the author, content title, publication name and Future Medicine Ltd as the original publisher.
6. The publisher will not amend, abridge, or otherwise change the content without authorization from Future Medicine Ltd.
7. Permission does not include any copyrighted material from other sources that may be incorporated within the content.
8. Failure to comply with the conditions above will result in immediate revocation of the permission here granted.

Date: 29/07/2013

OXFORD UNIVERSITY PRESS LICENSE
TERMS AND CONDITIONS

Jul 29, 2013

This is a License Agreement between Christiana A. Demetriou ("You") and Oxford University Press ("Oxford University Press") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Oxford University Press, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	3198150605145
License date	Jul 29, 2013
Licensed content publisher	Oxford University Press
Licensed content publication	International Journal of Epidemiology
Licensed content title	Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease:
Licensed content author	Caroline L Relton, George Davey Smith
Licensed content date	02/01/2012
Type of Use	Thesis/Dissertation
Institution name	
Title of your work	The impact of secular trends, dietary patterns and socioeconomic conditions on DNA methylation: an epigenome wide investigation
Publisher of your work	n/a
Expected publication date	Oct 2013
Permissions cost	0.00 EUR
Value added tax	0.00 EUR
Total	0.00 EUR
Total	0.00 EUR

Terms and Conditions

STANDARD TERMS AND CONDITIONS FOR REPRODUCTION OF MATERIAL FROM AN OXFORD UNIVERSITY PRESS JOURNAL

1. Use of the material is restricted to the type of use specified in your order details.

2. This permission covers the use of the material in the English language in the following territory: world. If you have requested additional permission to translate this material, the terms and conditions of this reuse will be set out in clause 12.

3. This permission is limited to the particular use authorized in (1) above and does not allow you to sanction its use elsewhere in any other format other than specified above, nor does it apply to quotations, images, artistic works etc that have been reproduced from other sources which may be

part of the material to be used.

4. No alteration, omission or addition is made to the material without our written consent. Permission must be re-cleared with Oxford University Press if/when you decide to reprint.

5. The following credit line appears wherever the material is used: author, title, journal, year, volume, issue number, pagination, by permission of Oxford University Press or the sponsoring society if the journal is a society journal. Where a journal is being published on behalf of a learned society, the details of that society must be included in the credit line.

6. For the reproduction of a full article from an Oxford University Press journal for whatever purpose, the corresponding author of the material concerned should be informed of the proposed use. Contact details for the corresponding authors of all Oxford University Press journal contact can be found alongside either the abstract or full text of the article concerned, accessible from www.oxfordjournals.org Should there be a problem clearing these rights, please contact journals.permissions@oxfordjournals.org

7. If the credit line or acknowledgement in our publication indicates that any of the figures, images or photos was reproduced, drawn or modified from an earlier source it will be necessary for you to clear this permission with the original publisher as well. If this permission has not been obtained, please note that this material cannot be included in your publication/photocopies.

8. While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by Oxford University Press or by Copyright Clearance Center (CCC)) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and Oxford University Press reserves the right to take any and all action to protect its copyright in the materials.

9. This license is personal to you and may not be sublicensed, assigned or transferred by you to any other person without Oxford University Press's written permission.

10. Oxford University Press reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

11. You hereby indemnify and agree to hold harmless Oxford University Press and CCC, and their respective officers, directors, employs and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

12. Other Terms and Conditions:

v1.4

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number RLNK501077270.
Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.

Make Payment To:
Copyright Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006

For suggestions or comments regarding this order, contact RightsLink Customer Support:customercare@copyright.com or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.

Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.
