Faculty of Pharmaceutical, Biomedical and Veterinary sciences
Department of Biomedical sciences
Protein chemistry, Proteomics and Epigenetic Signalling (PPES)

Epigenetic profiling of adverse lifestyle conditions and nutraceutical interventions in health and disease

Epigenetische profilering van ongunstige levensstijlfactoren en nutraceutische interventies bij gezondheid en ziekte

Thesis submitted for the degree of doctor in science: biochemistry and biotechnology at the University of Antwerp to be defended by Ken Declerck

Promotor: Prof. Dr. Wim Vanden Berghe

Antwerp, 2019

Members of the jury

Promotor:

Prof. Dr. Wim Vanden Berghe
Protein chemistry, Proteomics and Epigenetic Signalling (PPES), University of Antwerp, Belgium

Chairman:

Prof. Dr. Guy Van Camp
Center of Medical Genetics, University of Antwerp, Belgium

Internal members:

Prof. Dr. Wim Martinet
Physiopharmacology, University of Antwerp, Belgium

Prof. Dr. Stuart Maudsley
Translational Neurobiology, University of Antwerp, Belgium

External members:

Prof. Dr. Carmen Jerónimo
Cancer Biology and Epigenetics Group, IPO Porto Research Center (CI-IPOP), University of Porto, Portugal

Prof. Dr. Claudia Santos
Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Portugal
iBET, Instituto de Biologia Experimental e Tecnológica, Portugal

Table of contents

List of abbreviations

2-HG	2-hydroxyglutarate
3’MEC	3’-O-methyl(-)-epicatechin
4’MEC7G	4’-O-methyl(-)-epicatechin-7-β-D-glucuronide
5-AZA	5-azacytidine
5-Aza-CdR	5-aza-2’-deoxycytidine
5caC	5-carboxylcytosine
5fC	5-formylcytosine
5hmC	5-hydoxymethylcytosine
5mC	5-methylcytosine
5-MTHF	5-methyltetrahydrofolate
6mA	N6-methadenine
AD	Alzheimer's disease
AML	acute myeloid leukemia
aoSMC	aortic smooth muscle cells
AR	androgen receptor
AREase	arylesterase
athDMP	atherosclerosis-DMP
athDMR	atherosclerosis-DMR
Avy	viable yellow allele
AWHS	Aragon Workers' Health Study
AxinFu	axin fused allele
BEIP	Bucharest Early Intervention Project
BMIQ	beta mixture quantile dilation
CAUG	care as usual group
CB2	cannabinoid receptor 2
cfDNA	cell-free DNA
CGI	CpG island
chIP	chromatin immunoprecipitation
CML	chronic myelomonocytic leukemia
CRC	colorectal cancer
CRH	corticotropin-releasing hormone
cRNA	complementary RNA
CRP	C-reactive protein
CTCL	cutaneous T cell lymphoma
CTQ	childhood trauma questionnaire

CUP	carcinoma of unknown primary
CVD	cardiovascular disease
DEX	dexamethasone
DMP	differentially methylated position
DMR	differentially methylated region
DNMT	DNA methyltransferase
DNMTi	DNMT inhibitor
DOHaD	Developmental Origins of Health and Disease
EC4'S	(-)-epicatechin-4'-sulfate
EGCG	(-)-epigallocatechin-3-gallate
eQTL	expression quantitative trait locus
ERV	endogenous retrovirus
ET1	endothelin-1
EWAS	epigenome-wide association study
FACS	fluorescence-activated cell sorting
FCG	foster care group
FDR	false discovery rate
FH	fumarate dehydrogenase
FMD	flow-mediated vasodilation
gDMR	germline DMR
gDNA	genomic DNA
GEO	gene expression omnibus
GR	glucocorticoid receptor
GRE	glucocorticoid response elements
GWAS	genome-wide association study
HDACi	HDAC inhibitor
HOMA-IR	Homeostatic Model Assessment for Insulin Resistance
HP1	heterochromatin protein 1
HUVEC	human umbilical vein endothelial cells
IAP	intracisternal A-particle
IC	imprinted center
ICR	imprinted control region
IDH	isocitrate dehydrogenase
IFN	interferon
IHEC	International Human Epigenome Consortium
IPA	ingenuity pathway analysis
LAD	lamina-associated domain
LDTF	lineage-determining transcription factors

LOCK	large, organized chromatin lysine modifications
MAD	median absolute deviation
MBD	methyl-binding domain
mdNLR	methylation-derived neutrophil/lymphocyte ratio
MDS	myelodysplastic syndrome
meQTL	methylation quantitative trait loci
MOF	monomeric and oligomeric flavanols
MS	multiple sclerosis
MTHFR	methylenetetrahydrofolate reductase
NHP	non-human primate
NIG	never institutionalized group
NLR	neutrophil/lymphocyte ratio
NPC	neuronal precursor cells
PAI-1	plasminogen activator inhibitor type-1
PCA	principal component analysis
PD	Parkinson's disease
PGC	primordial germ cells
PGR	progesterone receptor
piRNA	PIWI-interacting RNA
ProgFib	ENCODE fibroblast
PTCL	peripheral T cell lymphoma
PTK	tyrosine kinase
PTSD	post-traumatic stress disorder
RCT	randomized controlled trial
SAH	S-adenosylhomocysteine
SAM	S-adenosyl methionine
SCLC	small-cell lung carcinoma
SDH	succinate dehydrogenase
SNP	single-nucleotide polymorphism
STK	serine/threonine kinase
TAD	topologically associated domains
TCGA	The Cancer Genome Atlas
TET	ten-eleven translocation proteins
TF	transcription factor
TFBS	transcription factor binding site
TSST	Trier Social Stress Test
UPD	uniparental disomy
VMP	variable methylated position

WGCNA	weighted correlation network analysis
WHO	World Health Organization

Summary

DNA methylation is an epigenetic mark which plays an important role in maintaining cellular identity by regulating cell and tissue-specific gene expression profiles. Epigenetic modifications are known to be influenced by both genetic and environmental factors and may therefore be seen as a bridge between the environment and the genome. Furthermore, epigenetic changes due to environmental stimuli may be mitotically inherited and thus also function as a cellular memory. Given these characteristics, DNA methylation is thought to mediate the long-term effects of chronic exposure to environmental factors on health and disease, which is also known as the Developmental Origins of Health and Disease (DOHaD) hypothesis. This paradigm proposes that the prenatal and early developmental stages are important contributors of an individual's health status throughout its life. Indeed, DNA methylation changes are now also identified in multiple lifestyle diseases, such as cancer, cardiovascular, metabolic and neurological diseases. Due to its dynamic nature, DNA methylation can be exploited as prognostic, diagnostic and therapeutic biomarkers, but also is an attractive target for therapeutic and nutritional interventions. In addition, studying epigenetic patterns could give us useful mechanistic insight in disease etiology and development. In this thesis, we studied DNA methylation patterns in easy-accessible tissues, including blood and saliva to monitor adverse early environmental exposures, cardio-metabolic diseases and nutritional intervention responses.

In the first part of the results section (chapters 3 and 4), we examine whether early adverse environmental conditions can be monitored through DNA methylation marks in blood and saliva. During prenatal and early development, the methylation profile is sensitive to adverse environmental factors and the exposure history builds up cellular memory which can promote diseases later in life. In this way, monitoring adverse DNA methylation changes holds promise to monitor and/or prevent potential disease risk later in life.

In chapter 3, we performed a pilot study in a cohort of Danish children, whose mothers were exposed to pesticides during pregnancy. In a previous study of those children, it was found that carriers of a polymorphism in the PON1 enzyme, which has an important role in organophosphate pesticide hydrolysis and is linked with atherosclerosis, have an adverse cardio-metabolic risk profile upon pesticide exposure. We could detect a specific DNA methylation profile in blood of prenatally pesticide exposed children carrying the *PON1* 192R-allele. Differentially methylated genes were enriched in several neuroendocrine signaling pathways including dopamine-DARPP32

feedback (appetite, reward pathways), corticotrophin releasing hormone signalling, nNOS, neuregulin signalling, mTOR signalling and type II diabetes mellitus signalling suggesting a possible link with the metabolic effects observed in these children. Furthermore, we were able to identify possible candidate genes which mediated the associations between pesticide exposure and increased leptin level, body fat percentage, and difference in BMI Z-score between birth and school age.
In chapter 4, we compared DNA methylation profiles in saliva samples of neglected orphan children of the well characterized Bucharest Early Intervention Project (BEIP), raised in institutions, following foster care or never-institutionalized children. We identified a set of 30 differentially methylated regions (DMRs) located in genes which are involved in nervous system and neuronal development, related to cognition, behavior and learning and psychological disorders. Further comparison with gene expression profiles before and after the Trier Social Stress Test (TSTT) identified a nuclear receptor coordinated stress dependent gene network, involving the glucocorticoid receptor. One of the genes with the highest correlation between TSST specific changes in cortisol levels, DNA methylation and gene expression was *CALD1*, a gene with established links to glucocorticoid stress responsiveness, neuronal migration and also the impact of cortisol stress hormones on neuronal morphology, dendritic spines and migration.

In the second part of the results section (chapter 5 and 6), we evaluated whether we could identify disease associated DNA methylation profiles in whole blood of atherosclerosis patients that could be used as a surrogate biomarker. Epigenetics is believed to play a significant role in the initiation and development of complex chronic diseases. For some diseases, the tissue of interest is difficult or even impossible to obtain, and therefore easily accessible surrogate tissues, like blood and saliva are recommended.
In chapter 5, we measured the genome-wide DNA methylation profile of whole blood samples of clinical atherosclerosis patients and compared it with epigenetic profiles of atherosclerotic plaque material, and subclinical atherosclerosis samples of the Aragon workers cohort, to identify potential surrogate markers for early cardiovascular disease (CVD) detection. We identified multiple DMRs in atherosclerosis patients related to epigenetic control of cell adhesion, chemotaxis, cytoskeletal reorganizations, cell proliferation, cell death, estrogen receptor pathways and phagocytic immune responses. Furthermore, a subset of 34 DMRs related to impaired oxidative stress, DNA repair, and inflammatory pathways could be replicated in an independent cohort study of donor-matched healthy and atherosclerotic human aorta tissue and human carotid plaque samples. Upon integrated network analysis, BRCA1 and CRISP2 DMRs were

identified as most central disease-associated DNA methylation biomarkers. Differentially methylated BRCA1 and CRISP2 regions could be further replicated in blood, aorta tissue and carotid plaque material of atherosclerosis patients. Moreover, methylation changes at BRCA1 and CRISP2 specific CpG sites were consistently associated with subclinical atherosclerosis measures (coronary calcium score and carotid intima media thickness) in an independent sample cohort of middle-aged men with subclinical CVD in the Aragon Workers' Health Study (AWHS). Therefore, BRCA1 and CRISP2 DMRs hold promise as novel blood surrogate markers for early risk stratification and CVD prevention.

In chapter 6, we cross-compared the atherosclerosis specific methylation profile identified in chapter 5 with whole blood DNA methylation profiles of Alzheimer's disease (AD) patients to identify a common DNA methylation profile which may be valuable as a blood-based DNA methylation inflammaging biomarker. Using publicly available 450k Illumina methylation datasets, we identified a co-methylation network associated with both atherosclerosis and AD in whole blood samples. This methylation profile appeared to indicate shifts in blood immune cell type distribution. Remarkably, similar methylation changes were also detected in disease tissues, including AD brain tissues, atherosclerotic plaques and tumors and were found to correlate with immune cell infiltration. In addition, this immune-related methylation profile could also be detected in other inflammaging diseases, including Parkinson's disease and obesity, but not in multiple sclerosis, schizophrenia and osteoporosis. In addition to epigenetic clock measurements, this immune-methylation signature may become a valuable blood-based biomarker of resilient immune fitness to prevent chronic inflammatory disease development or monitor lifestyle intervention strategies which promote healthy aging.

In the final part of the results section (chapter 7 and 8), we studied the epigenetic effects of nutritional and immunomodulatory phytochemical compounds on DNA methylation. Moreover, we addressed to which extent disease associated DNA methylation patterns can be reversed by nutraceutical and/or phytomedicinal specific interventions.

In chapter 7, we examined DNA methylation changes upon *in vitro* exposure of endothelial cells to cardio-protective flavanols and in blood samples following an eight-week diet intervention with flavanol-rich grape seed extract. Diets rich in epicatechin flavanols are known to exert cardioprotective effects through reduction of monocyte-endothelial cell adhesion and transendothelial monocyte migration. In line with their known biological properties, flavanol metabolites revealed endothelial DNA methylation changes in genes involved in cell adhesion, cytoskeletal organization, renin angiotensin, nitric oxide and axonal guidance signaling pathways. Upon flavanol

enriched diet intervention, we observed a strong interindividual variation in blood DNA gene methylation response which affects common cell adhesion pathways. The high interindividual gene diversity suggests highly redundant epigenetic wiring of cell adhesion networks, beyond direct transcriptional regulation. Furthermore, we identified reciprocal atherosclerosis and flavanol diet specific epigenetic changes, which involve systemic immunological changes in blood cell types.
In chapter 8, we studied epigenetic changes upon *in vitro* exposure of THP-1 monocytes to Echinaforce®, a commercial standardized immunomodulatory *Echinacea purpurea* extract, which can protect against colds and respiratory infections. Upon integrating transcriptome, kinome and DNA-methylome profiles of THP-1 monocyte cells treated with Echinaforce®, we observed activation of antiviral immunological pathways. More particularly, our systems biology approach demonstrates that Echinaforce® treatment of THP1 cells triggers interferon and NF-κB kinase signaling pathways which promote innate immunity gene expression changes, macrophage M1 polarization and DNA hypermethylation of repeat elements in CpG poor gene bodies and intergenic regions.

The results of this thesis clearly demonstrate that DNA methylation profiles in blood and/or saliva are associated with environmental exposures and chronic lifestyle diseases, and that nutritional and immunomodulatory phytochemical compounds are able to modulate DNA methylation. However, to fully exploit the potential of epigenetics in (personalized) health applications, multiple hurdles still need to be overcome. For instance, which cells/tissues are most sensitive or predictive for adverse health conditions? For which diseases we can use blood or saliva based surrogate biomarkers in epigenetic-based health prediction? How to interpret biological impact of DNA methylation changes in relation to gene expression, genome stability or evolutionary drift? Are they a cause or consequence of the disease process/environmental exposure? These problems, together with recommendations for future epigenetic studies, will be discussed in the final chapter of this thesis.

Samenvatting

DNA-methylatie is een epigenetische markering die een belangrijke rol speelt bij het behouden van de cellulaire identiteit door regulatie van cel en weefsel-specifieke genexpressie profielen. Epigenetische modificaties worden beïnvloed door zowel genetische als omgevingsfactoren en kunnen daardoor gezien worden als de brug tussen omgeving en genoom. Bovendien worden omgevings-geïnduceerde epigenetische veranderingen elke celdeling overgeërfd, waardoor epigenetische modificaties ook kunnen fungeren als cellulair geheugen. Deze eigenschappen maken van DNA-methylatie een mogelijke kandidaat om de langetermijn effecten van omgevingsfactoren op de gezondheidstatus te verklaren, welke ook gekend is als de *Developmental Origin of Health and Disease (DOHaD)* hypothese. Volgens deze hypothese zijn de prenatale en vroege ontwikkelingsstadia sterk bepalend voor de gezondheidsstatus van een individu op volwassen leeftijd. DNA-methylatie veranderingen zijn inderdaad nu ook geïdentificeerd in verschillende levensstijlziekten, zoals, kanker, cardiovasculaire, metabolische en neurologische ziekten. Door zijn reversibel karakter zou DNA-methylatie informatie gebruikt kunnen worden als prognostische, diagnostische en therapeutische biomerker, maar zou het ook een mogelijke doelwit kunnen zijn voor therapeutische en nutritionele interventies. Bovendien kunnen epigenetische patronen ons belangrijke informatie verschaffen in de etiologie en ontwikkeling van ziekten. In deze thesis hebben we DNA-methylatie patronen bestudeerd in makkelijk toegankelijke weefsels, zoals bloed en speeksel om vroege nadelige blootstellingen aan omgevingsfactoren, cardio-metabole ziekten en nutritionele interventie responsen op te volgen.

In het eerste deel van de resultaten (hoofdstuk 3 en 4) gaan we na of we vroege nadelige omgevingsfactoren kunnen detecteren via DNA-methylatiemerkers in bloed en speeksel. Het methylatieprofiel is voornamelijk gevoelig aan nadelige omgevingsfactoren tijdens de prenatale en vroege ontwikkeling waarbij gradueel een cellulair geheugen wordt opgebouwd dat kan bijdragen tot ziekten later in het leven. Vroegtijdige detectie van mogelijk nadelige DNA-methylatieveranderingen zou dus zowel gebruikt kunnen worden om ontstaan van ziekten op te sporen als te voorkomen. In hoofdstuk 3 werd een pilootstudie uitgevoerd in een cohort van Deense kinderen waarvan de moeders blootgesteld werden aan pesticides tijdens de zwangerschap. In een vorige studie in dezelfde kinderen, vond men dat dragers van een polymorfisme in het PON1 enzym, welke een belangrijke rol speelt in de hydrolyse van organofosfaat pesticides en gelinkt is met atherosclerose, een nadelig effect heeft op het cardio-metabolisch risicoprofiel na pesticideblootstelling. We konden een specifiek DNA-

methylatieprofiel detecteren in bloed van de kinderen die prenataal blootgesteld werden aan de pesticides en drager zijn van het PON1 192R-allel. Differentieel gemethyleerde genen waren aangerijkt in verschillende neuroendocriene signaalwegen zoals dopamine-DARPP32 terugkoppeling (eetlust, beloning signaalwegen), corticotropin-releasing hormoon, nNOS, neureguline signalering, mTOR signalering en type 2 diabetes mellitus signaalweg wat wijst op een mogelijke verband met de metabolische effecten in deze kinderen. Bovendien, konden we een aantal mogelijke kandidaatgenen identificeren die de associatie tussen pesticide blootstelling en verhoogde leptine levels, lichaamsvet percentage en verschil in BMI Z-score tussen geboorte en school leeftijd mediëren.
In hoofdstuk 4 vergeleken we de DNA-methylatie profielen in speekselstalen van verwaarloosde weeskinderen van de goed gekarakteriseerde Bucharest Early Intervention Project (BEIP), opgegroeid in instituten voor weeskinderen, pleegzorg of nooit-geïnstitutionaliseerde kinderen. We identificeerden een set van 30 differentieel gemethyleerde regio's in genen welke betrokken zijn in de ontwikkeling van het zenuwstelsel en neuronen, en gerelateerd aan geheugen, gedrag en leren, en psychologische aandoeningen. Verdere vergelijking met genexpressie-profielen voor en na de Trier Social Stress Test (TSTT) identificeerde een stress hormoonreceptor gecoördineerd netwerk, waarbij de glucocorticoïde receptor betrokken is. Eén van de genen met een sterke correlatie tussen TSST-specifieke veranderingen in cortisol niveaus, DNA-methylatie en genexpressie was *CALD1*, een gen betrokken bij glucocorticoïde stress respons, neuronale migratie en die ook de morfologie van neuronen, dendritische spines en migratie beïnvloedt.

In het tweede deel van de resultaten (hoofdstuk 5 en 6) gingen we na of bloedstalen bruikbaar zijn voor epigenetische diagnostiek bij atherosclerose patiënten, daar plaque materiaal onmogelijk te verkrijgen is in routine-onderzoek en men dus aangewezen is op makkelijk beschikbare surrogaat weefsels, zoals bloed en speeksel.
In hoofdstuk 5 hebben we het genoomwijde DNA-methylatie profiel van bloedstalen van klinische atherosclerose patiënten bepaald en vergeleken met de epigenetische profielen van atherosclerotische plaques, en subklinische atherosclerose stalen afkomstige van een Aragon werkers cohort, om zo mogelijke surrogaat merkers voor vroege cardiovasculaire aandoeningen te detecteren. In atherosclerosis patiënten konden verschillende differentieel gemethyleerde regio's (DMRs) worden geïdentificeerd die betrokken zijn in epigenetische controle van celadhesie, chemotaxis, cytoskelet reorganisatie, cel proliferatie, celdood, estrogeen receptor signaalwegen en fagocytotische immuun-responsen. Bovendien, kon ook een subset van 34 DMRs betrokken bij oxidatieve stress, DNA-herstel en inflammatie signaalwegen

gerepliceerd worden in een onafhankelijke cohort van donor-gepaarde gezonde en atherosclerotische humane aorta-weefsels en humane carotide plaque stalen. Na geïntegreerde netwerk analyse werden BRCA1 en CRISP2 DMRs geïdentificeerd en gevalideerd als de meest centrale ziekte-geassocieerde DNA-methylatie biomerkers. Bovendien waren de methylatie-veranderingen in specifieke BRCA1 en CRISP2 CpG sites consistent geassocieerd met subklinische atherosclerose merkers (coronaire calcium score en carotide intima-media-dikte) in een onafhankelijke cohort van mannen van middelbare leeftijd in de Aragon Worker's Health studie. BRCA1 en CRISP2 DMRs zijn dus veelbelovende nieuwe bloed surrogaat merkers voor vroeg-risico stratificatie en preventie van cardiovasculaire aandoeningen.
In hoofdstuk 6 werd het atherosclerose specifiek profiel dat geïdentificeerd werd in hoofdstuk 5, vergeleken met bloed DNA-methylatie profielen van Alzheimer patiënten om op die manier een gemeenschappelijk DNA methylatie profiel te identificeren dat kan dienen als bloed-gebaseerde DNA-methylatie biomerker voor *inflammaging*. Door gebruik te maken van publiek beschikbare 450k Illumina methylatie datasets, konden we een co-methylatie netwerk identificeren dat zowel geassocieerd is met atherosclerose als met de ziekte van Alzheimer in bloed. Dit methylatieprofiel blijkt het resultaat te zijn van veranderingen in de compositie van bloed immuun-celtypes. Interessant is dat vergelijkbare methylatieveranderingen ook gedetecteerd konden worden in weefsels, zoals hersenweefsels van Alzheimer patiënten, atherosclerotische plaques en tumoren, en gecorreleerd was met immuuncel infiltratie. Bovendien kon dit immuun-gerelateerde methylatiepatroon ook gedetecteerd worden in ander *inflammaging* ziekten, zoals de ziekte van Parkinson en obesitas, maar niet in multipel sclerose, schizofrenie en osteoporose. Complementair aan de epigenetische klok kan dit immuun-methylatie profiel een waardevolle bloed-gebaseerde biomerker worden voor immunologische fitheid en veerkracht om chronische ziekten te voorkomen of levensstijl interventies te begeleiden voor gezonde veroudering.

In het laatste deel van de resultaten (hoofdstuk 7 en 8) werden de epigenetische effecten van nutritionele en immuno-modulerende fytochemicaliën op DNA-methylatie bestudeerd. Bovendien, werd er ook nagegaan in hoeverre ziekte-geassocieerde DNA-methylatie patronen konden omgekeerd worden door nutraceutische en/of fytomedicinale specifieke interventies.
In hoofdstuk 7 onderzochten we DNA-methylatieveranderingen na *in vitro* behandeling van endotheelcellen met cardio-protectieve flavanolen en in bloedstalen na een acht-weken durende dieetinterventie met een flavanol-rijk druivenpit-extract. Voeding aangerijkt met epicatechine flavanolen werkt cardioprotectief door het verzwakken van monocyt-endotheel celadhesie en transendothele monocyt migratie. In lijn met

deze gekende biologische eigenschappen, werden DNA-methylatieveranderingen waargenomen in genen die betrokken zijn bij celadhesie, cytoskelet organisatie, renin-angiotensine, stikstof-oxide en axonale geleiding signaalwegen. Na een flavanol-rijke dieetinterventie, konden we sterke interindividuele variatie oberveren in DNA-methylatie veranderingen welke gemeenschappelijk celadhesie signaalwegen beïnvloeden. Deze sterke interindividuele gendiversiteit suggereert sterk redundante epigenetische modulatie van cel-adhesie signaalnetwerken, zonder sterke effecten op genexpressie-regulatie. Bovendien konden we tegengestelde epigenetische veranderingen aantonen bij atherosclerose en flavanol-dieet interventie waarbij veranderingen in samenstelling van bloed celtypes betrokken zijn.
In hoofdstuk 8, onderzochten we epigenetische veranderingen na *in vitro* behandeling van THP-1 monocyten met Echinaforce®, een commercieel gestandaardiseerd immuno-modulerend *Echinacea purpurea* extract, dat bescherming biedt tegen verkoudheden en respiratoire infecties. Na geïntegreerde transcriptoom, kinoom en DNA methyloom profilering van THP-1 monocyten behandeld met Echinaforce®, konden we de activatie van antivirale immunologische signaalwegen detecteren. Meer specifiek toonde onze systeem biologische analyse aan dat Echinaforce® behandeling interferon en NF-κB kinase signaalwegen aanslaat die genexpressie van het aangeboren immuunsysteem, macrofaag M1 polarisatie en DNA hypermethylatie van repeat elementen in CpG-arme gen en intergenische regio's stimuleren.

Tot slot tonen de resultaten van deze thesis aan dat DNA-methylatie profielen in bloed en/of speeksel geassocieerd zijn met blootstellingen aan omgevingsfactoren en chronische levensstijl ziekten, die kunnen beïnvloed worden door nutritionele en immuno-modulerende fytochemicaliën. Om epigenetische biomerkers diagnostisch in te zetten bij (gepersonaliseerde) gezondheidstoepassingen moeten er echter nog verschillende obstakels overwonnen worden. Zoals bijvoorbeeld, welke cellen/weefsels zijn meest gevoelig of betrouwbaar voor het meten van schadelijke blootstelling? Voor welke ziektes kan bloed en speeksel gebruikt worden als surrogaat biomerker voor epigenetisch gebaseerde gezondheidspredictie? Hoe interpreteren we biologische impact van DNA-methylatie veranderingen op genexpressie, genoomstabliteit of evolutionaire drift? Zijn ze een oorzaak of gevolg van het ziekteproces en/of omgevingsblootstelling? Deze problemen, samen met aanbevelingen voor toekomstige epigenetische studies zullen besproken worden in het laatste hoofdstuk van deze thesis.

1

Introduction

Most human traits and common diseases are influenced by a complex interplay of genetic, environmental and lifestyle factors. In 2001, the first sequence of the human genome was published [1]. Since then, genome sequencing contributed tremendously in the understanding of the genetic and molecular basis of different complex human diseases, including diabetes type II, obesity, CVDs, Alzheimer's disease, asthma, and many more. Genome-wide association studies (GWAS) identified many genetic variants which may predispose an individual to developing a certain disease [2]. Furthermore, genetic biomarkers may also help in predicting drug response and toxicity, and are therefore promising in personalized medicine [3]. Of interest, a lot of these markers are now being used in commercially available tests to inform healthy individuals about their genetic risk profile. However, it also became clear that genetics couldn't explain the complete heritability of complex diseases and that most of the identified single-nucleotide polymorphisms (SNPs) only make small contributions to the disease risk [4]. Environmental and lifestyle factors are playing an at least equally important role in human disease and health, and are known to interact with genetic factors. For example, it is known that environmental conditions during pregnancy may have a strong impact on the health of the offspring later in life [5]. Therefore, heritable risk for certain complex diseases may not only depend on the parents genetic make-up but also on environmental conditions during intra-uterine and even postnatal early development, which is known as the Developmental Origins of Health and Disease (DOHaD) hypothesis [5-7].

How early environmental factors may impact the health of the offspring is not completely understood, but one promising mechanism is through influencing a person's epigenetic profile or epigenome [6, 7]. The main function of epigenetic factors is maintaining the cell's transcriptional program and cellular identity after every cell division [8]. Importantly, epigenetic factors are being influenced by both genetic and environmental factors and may therefore be seen as a bridge between the environment and the genome [9]. Furthermore, epigenetic changes due to environmental stimuli may be mitotically inherited and thus also function as a cellular memory [10, 11].

This thesis is focused on epigenetic mechanisms, in particularly DNA methylation, and the use of this information to explain environment-phenotype interactions and to develop biomarkers. Here, I will introduce the concept of epigenetics, why it is important, how it may contribute to disease and health, and how it can be used in biomedical applications.

Epigenetics as a regulator of cellular identity

The human body contains 37 trillion cells classified into around 200 different cell types. Although each cell is originally derived from one stem cell and contains the same genetic material, different cells can differ tremendously in shape and function. These phenotypic differences across cell types are mainly the result of differences in the cell's transcriptional profile. A neuron for example, mainly expresses genes which have a role in transmitting electrical and chemical signals to other cells. In contrast, leukocytes express mainly genes important in immune function to protect the body from bacteria and viruses. Therefore, the cell needs a system that controls the activity of genes specific for each cell type and which maintains this transcriptional profile as the cell divides. The main players in this system are transcription factors (TF) and epigenetic modifications.

One of the most important classes of proteins in the control of gene expression are TF. They bind DNA at enhancer and promoter regions of genes and recruit cofactors and RNA polymerase II necessary to initiate and elongate gene transcription [12, 13]. TFs bind typically on specific DNA recognition motifs. In addition, the binding of TF is also influenced by how dense the DNA is packed [14]. DNA can be packed in an open 'euchromatin' state or in a closed 'heterochromatin' state. In the open euchromatin conformation the DNA strand is accessible for the transcription apparatus, however in the closed conformation the dense DNA packaging prevents TFs from binding to the DNA and initiating gene expression. It is at this stage that epigenetic mechanisms play an important role.

Genomic DNA is packed in the nucleus by histone proteins into, so called, nucleosomes. A nucleosome consists of 147 bp DNA strand wrapped around an octamer of histones consisting of two copies of each of the core histones (H2A, H2B, H3 and H4). The N-terminal tails of these histones are extensively modified by post-translational modifications (acetylation, methylation, phosphorylation, …). These histone modifications or marks represent a first form of epigenetic information (**Figure 1**) [15]. They provide docking sites for the so called histone readers which are by itself transcriptional regulators or recruit other co-factors leading to changes in nucleosome occupancy and gene expression. Interestingly, specific histone marks are associated with different transcriptional states. For example, histone acetylation is associated with an open euchromatin conformation of the DNA strand and therefore permits gene activation, while the tri-methylation of lysine 9 of histone H3 (H3K9me3) is associated with the closed transcriptional repressed heterochromatin state. Similar, H3K4

methylation is mainly associated with transcriptional activation: H3K4me3 is characteristic for active promoters, while H3K4me1 is found in active enhancers. In this way the genome can be divided into a relative small subset of chromatin states, including active/repressed promoters, strong/weak enhancers, heterochromatin, transcriptional elongation regions, etc. Histone modifications can be added and removed by the so called histone writers and histone erasers, respectively.

Another important epigenetic mechanism is DNA methylation [11], which will be the focus of this thesis. During DNA methylation, a methyl group is added to a cytosine nucleotide leading to the formation of 5-methylcytosine (5mC). In humans and most other animals, methylation is almost exclusively found on cytosine nucleotides followed by a guanine nucleotide (the so called CpG-sites). However, also non-CpG methylation (CpA, CpT and CpC) exists but are far less abundant than CpG methylation. Non-CpG methylation is only detected in neuronal cells [16] and embryonic stem cells [17] (in contrast, plants have a much higher amount of non-CpG methylation). Although the function of human non-CpG methylation is further emerging, in this thesis the focus is only on CpG methylation. In general, DNA methylation at gene promoter regions leads to the repression of gene expression, probably by preventing TFs from binding and/or by the binding of methyl-binding domain (MBD) proteins recruiting other transcriptional regulators. More details about the mechanism, function and applications of DNA methylation will be given in the next sections.

Beside histone modifications and DNA methylation, other epigenetic mechanisms have been described. For example non-coding RNAs may also be seen as a contributor of epigenetics and collaborate together with DNA methylation in the control of gene imprinting and X-chromosome inactivation [19]. Another form of epigenetic mechanism is the higher-order chromatin structure, including large, organized chromatin lysine modifications (LOCKs), lamina-associated domains (LADs) or topologically associated domains (TADs) [20-22].

Histone modifications and DNA methylation profiles are specific for each cell type, and therefore contribute to the cellular identity of the cell. While epigenetic mechanisms rather act as stabilizers and maintainers of cellular states, TFs can be seen as specifiers of cellular identity. Cell-type specific enhancers are selected and activated by pioneer or lineage-determining transcription factors (LDTF) [23, 24]. They are able to bind nucleosome-compacted DNA and may open the chromatin conformation thereby driving lineage-specific transcription programs. Interestingly, cells can be reprogrammed *in vitro* to a pluripotent state by the ectopic expression of only four TFs: OCT4, SOX2, KLF4 and MYC [25]. In addition, cells can also be directly reprogrammed towards specific cell lineages [26]. However, the efficiency of these methods are limited and are mainly hampered by epigenetic barriers [27-31]. Removing these epigenetic

barriers may improve reprogramming efficiency [27]. Therefore it seems that both TFs and chromatin structure contribute to cellular states and that the relationship between TFs and chromatin structure is bidirectional: TF binding can induce local epigenetic and chromatin changes, while epigenetic modifications and chromatin structure may prevent TFs from binding.

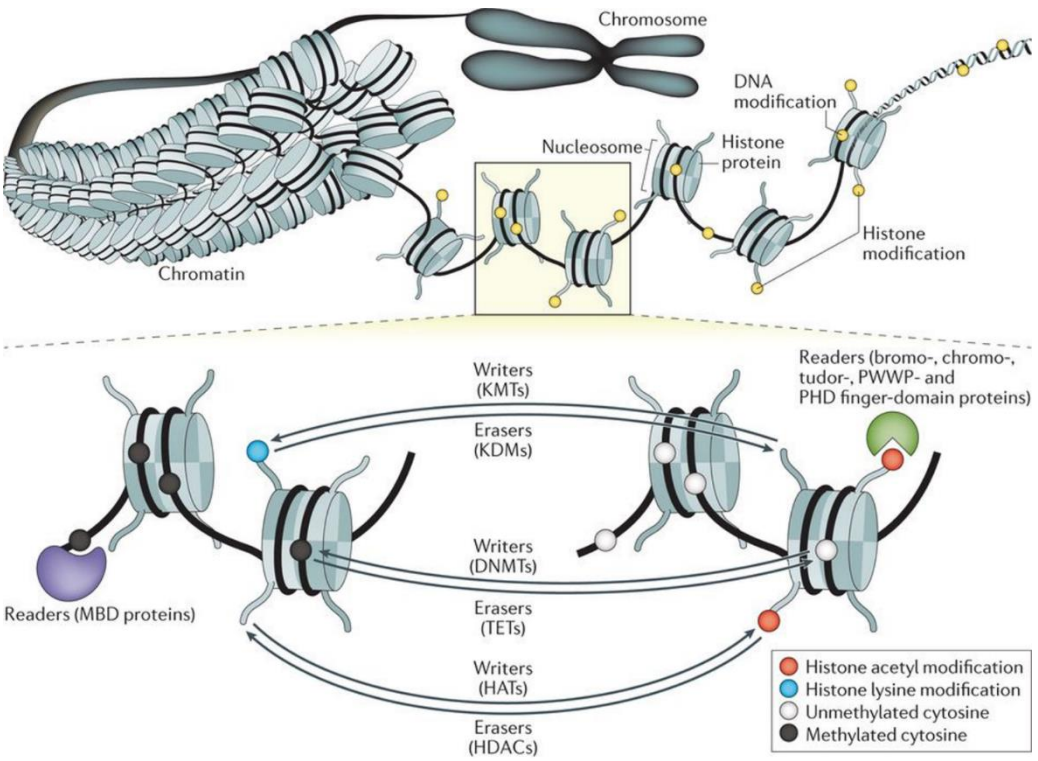


Figure 1: Major epigenetic mechanisms. DNA is packed into chromatin consisting of DNA wrapped around histone proteins, which is called a nucleosome. Histones can be post-translationally modified by specific histone writers (KMTs, HATs, etc) and erasers (KDMs, HDACs, etc). Methyl groups are added to DNA by DNMTs and removed by TET enzymes. Both histone modifications and DNA methylation are docking sites for specific readers. KMT: Lysine methyltransferase; KDM: Lysin demethylase; HAT: Histone acetyltransferase; HDAC: Histone deacetylase; DNMT: DNA methyltransferase; TET: Ten-eleven translocation enzymes. Figure was adapted from [18].

DNA methylation patterns and landscapes

The human genome contains around 30 million CpG sites which is less what would be expected and is mainly due to the spontaneously and enzymatically deamination of

5mC to thymine [32]. Interestingly, a C to T transition at CpG dinucleotides is the most frequent mutation observed in human diseases [33, 34]. CpG sites are not equally distributed and tend to cluster in so called CpG islands (CGI), which are regions of around 1 kb with a high density of CpG sites compared to the regions outside CGIs (in fact the CG content in CGIs are comparable to the expected CG content) [35, 36]. About 50% of the CGIs are located at promoter sites of known annotated genes. Furthermore, about 70% of annotated promoters possess a CGI, especially at promoters of housekeeping and developmental regulator genes. The other half of the CGIs (which are sometimes called orphan CGIs) are either located in gene bodies or between genes in almost an equal proportion [35, 36].

CGIs are mainly unmethylated, while CpGs outside these islands are highly methylated. Because CGIs are almost never methylated in the germline, they are therefore less vulnerable to the spontaneous deamination of methylated cytosines and protected from the loss of CpG sites. It has been shown that methylation at CGIs is rather stable across different cell types, and that regions just next to these islands, CGI shores, are more variably methylated across tissues, and better associated with gene expression [37].

Thus, genomic DNA methylation profile (methylome) follows a bimodal distribution, with unmethylated CpGs (<10%) clustering in CGIs, and highly methylated CpGs (>85%) located in CpG-poor regions outside CGIs (**Figure 2**).

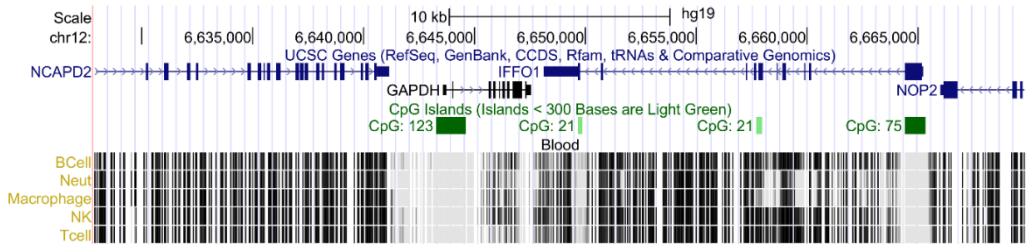


Figure 2: Distribution of DNA methylation. A genomic region is shown around the human GAPDH gene. A CGI is found at the promoter region of GAPDH (CpG: 123). Also the promoter region of IFFO1 contains a CGI (CpG: 75). Below, the methylation levels are shown from five human blood cell types demonstrating low methylation (light gray) in CGI, and high methylation (black) outside these islands. Also note, the intragenic CGI in the IFFO1 gene (CpG: 21), where methylation is much more variable across the five blood cell types. Figure was created using the UCSC genome browser.

Establishing and maintaining methylation patterns

During embryonic development, the methylome undergoes very dynamic DNA demethylation and remethylation phases (**Figure 3**) [38, 39]. After fertilization, totipotency of the cell is re-acquisitioned by the (almost) complete erasure of paternal and maternal genome. DNA methylation patterns are subsequently re-established by

de novo methylation at about the stage of implantation. Another demethylation phase occurs during the formation of gametes. These rapid de- and re-methylation reactions are being catalyzed by ten-eleven translocation (TET) family of proteins and DNA methyltransferases (DNMTs), respectively. S-adenosyl methionine (SAM) serves as the methyl donor during the methylation reaction.

After fertilization, the global DNA demethylation is different between the paternal and maternal genomes. It is believed that the demethylation of the maternal genome is mainly accomplished by a passive demethylation process: multiple rounds of DNA replication without DNA methylation maintenance leads to the dilution of 5mC [38, 40]. In contrast, the demethylation of the paternal genome occurs much faster and starts before the first replication round and therefore cannot only be explained by this passive process [41]. Here, TET enzymes helps in catalyzing an active form of demethylation [42]. TET3 converts methylated cytosine into 5-hydroxymethylcytosine (5hmC) [43, 44], which are subsequently further oxidized to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). Demethylation is then established by either the dilution of these oxidized 5mC modification after multiple rounds of DNA replication [45, 46] or excision of 5fC and 5caC by thymine-DNA glycosylases followed by base excision repair [42]. Recent research however, showed that TET3 is also involved in the demethylation of maternal genome [47], and that both paternal and maternal genomes undergo active and passive demethylation [48, 49]. It was shown that for both genomes a passive replication-dependent dilution was the main way for global demethylation [47, 50]. Furthermore, as will be discussed later, this demethylation phase is not complete and some regions, including imprinted centers (IC), may escape from erasure.

After erasure in the early embryo (blastocyst stage), upregulation of the *de novo* DNMTs, DNMT3A and -3B leads to the re-establishment of the DNA methylation pattern after implantation [51]. Another protein DNMT3L, which lacks the methyltransferase catalytic domain, is crucial in the *de novo* methylation by interacting with unmethylated H3K4 and recruiting of DNMT3 [52, 53]. Furthermore, it can enhance the activity of DNMT3 [54-56]. CGIs are protected from *de novo* methylation during this remethylation phase, probably due to the binding of RNA polymerase and other TFs [57, 58], however, the precise mechanism is not completely understood. Also histone modifications may play a role in this protection. For example, H3K4me3 histone mark, which is often associated with transcription start sites, allosterically inhibits the binding of *de novo* methylases [52, 53]. Furthermore, multiple proteins, harboring a CXXC domain (CFP1, MLL1 and -2, KDM2A and -B, TET1 and -3), bind to unmethylated CpGs which may subsequently prevent aberrant *de novo* methylation by recruiting histone methyltransferases and demethylases and removing unwanted DNA

methylation marks [59, 60]. For example, CFP1 recruits H3K4 methyltransferases ensuring the exclusion of DNMTs [61, 62].

At the end of the re-establishment phase, *de novo* methylases are downregulated and the maintenance methylase, DNMT1, which is constantly associated with the DNA replication machinery, ensures the inheritance of this initial pattern after every cell division [63]. DNMT1 has a preference of methylating hemi-methylated DNA strands, and therefore uses the methylation marks from the parental DNA strand as a template to copy these onto the newly synthesized daughter strands. Other factors, including UHRF1 help in recruiting DNMT1 to hemi-methylated DNA [64, 65]. Although, in the classical view, DNMT1 was seen as the maintenance methylase and DNMT3 as *de novo* methylase, further studies showed that DNMT1 also has *de novo* methylation activity [66, 67] and that DNMT3 is also needed in the maintenance of DNA methylation [68, 69]. Therefore, a more complex stochastic model of DNA methylation was proposed, where DNA methylation at each site is determined by the local rates of DNA methylation and demethylation [68]. In addition, other mechanisms are needed to maintain cellular DNA methylation patterns, including TF binding and histone modifications [70-72], which protect regions from methylation or target regions for methylation.

During cellular differentiation, around 20% of all CpGs undergoes dynamic DNA methylation changes in a sequence-specific way resulting in cell-specific DNA methylation profiles [73]. For example, many genes needed for pluripotency in the embryonic stem cells, including Oct3/4 and Nanog, are silenced by promoter hypermethylation during differentiation [74]. Other regions that become *de novo* methylated are mainly repeat elements, including satellite DNA (pericentromeric repeats) and transposons [39]. Interestingly, *de novo* methylation is almost always mediated by histone methylases that are recruited by local regulatory factors: H3K9 methylase G9a for promoter methylation, H3K9 methylase Suv39v for satellite DNA (pericentromeric repeats), and H3K9 methylases G9a and SETDB1 for transposable elements [75]. This also implies that DNA methylation is not an initiator of promoter silencing, but rather acts as a secondary mechanism ensuring long-term maintenance of the repressed state [76]. Targeted demethylation occurs in genes which are expressed in a tissue-specific way. TET1, 2 and -3 enzymes are responsible for this demethylation and are required for proper differentiation as indicated by TET knockouts cells [77-81]. Similar as with *de novo* methylation, trans-acting factors (e.g. TFs) binding to cis-acting sequences are often necessary to induce demethylation of that region [80, 82, 83]. Furthermore, there exists a correlation between tissue-specific demethylation and TF binding at the same sites [84].

Cellular differentiation leads to cell type-specific DNA methylomes. Studies which compared methylation patterns across different cell and tissue types demonstrated that most of the methylation differences are not located in CGI promoter sites, but rather are enriched in CpG-poor regions [73, 85, 86]. In addition, these tissue-associated differentially methylated regions (DMRs) were predominantly mapped to distal DNase hypersensitive sites, TF binding sites and cell-type specific enhancers. Among CGI, gene body CGIs are more variable as compared to promoter CGIs which points towards a function in alternative transcription [87, 88].

A second global demethylation phase occurs during the specification of the germ line and the formation of primordial germ cells (PGCs) leading to the expression of germ-line specific genes [89]. During gamete formation also the imprinted regions are now completely demethylated, however, still some regions may escape from this complete erasure, opening a window for trans- and intergeneration epigenetic inheritance [90]. Global demethylation in PGCs occurs in a sequential manner [91-93] with a first passive demethylation phase due to the repression of UHRF1, DNMT3A and -B [94], and a subsequently more locus-specific demethylation phase due to the action of TET1 and -2, and the formation of 5-hydoxymethylcytosine [95, 96].

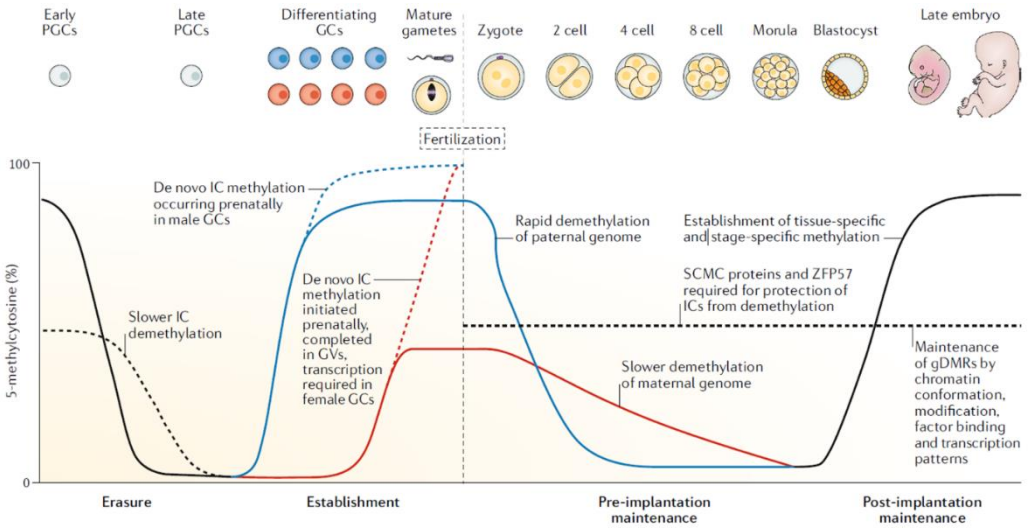


Figure 3: DNA methylation dynamics during human development. DNA methylation is erased in PGCs and re-established during final gamete formation in a sex-specific way. During gamete formation also methylation in imprinted centers (IC, dashed lines) are re-established. Fertilization leads again to a global demethylation with different kinetics between maternal (red line) and paternal genome (blue line). ICs are not erased (black dashed line). DNA methylation is re-established after implantation. Figure is adapted from [102].

The remethylation phase of the gametes is sex-specific; whereas in male gametes, the genome is remethylated before birth, the female gametic genome retains its complete

methylation pattern until sexual maturation. Remethylation is mainly established by DNMT3A and DNTM3L [97, 98]. However, also other factors, including histone modifications [99, 100] and PIWI-interacting RNAs (piRNA) [101] were shown to be essential in *de novo* DNA methylation in gametes. In this way, genomic imprints are again being established.

Shaping the human methylome: genetic and environmental influences

DNA methylation patterns not only differ across different cell- and tissue types but are also variable (although to a lesser extent) across different individuals. Whole-genome bisulfite sequencing showed that approximately 15-25% of all the CpGs are variable [103]. Similar as with tissue type-associated differentially methylated sites, the most variable CpG sites across individuals are enriched in enhancers and depleted in promoters [103-106]. Despite some CpG sites showing strong variability, the majority of CpG sites remain rather stable with either low methylation values (~0%) or high methylation values (~100%) [103]. Both genetic, stochastic, environmental and lifestyle factors have been shown to impact DNA methylation variability (**Figure 4**).

Genetic factors. Studies in twins have been valuable in dissecting genetic and non-genetic influences on DNA methylation and to estimate the heritability of DNA methylation patterns [107]. Multiple studies demonstrated a higher correlation in DNA methylation between monozygotic twins as compared to dizygotic twins, indicating a genetic contribution to DNA methylation [107-109]. The average heritability of DNA methylation was estimated to be rather low, around 20%, but can differ tremendously among CpG sites, with some sites having a heritability towards 100% [106, 110-112]. In addition, this estimated heritability was mainly determined by array-based approaches, which primary covers CGI and promoter regions that are known to be relatively stable, and therefore less affected by genetics. Indeed, when variance was taken into account, the average heritability increases towards 35% [111].

The genetic contribution of DNA methylation variability has been further supported by studies which found associations between genetic variants (SNPs) and DNA methylation at specific CpG sites in different tissues [113, 114]. These SNPs are called methylation quantitative trait loci (meQTLs). In general, the associated CpG sites display a characteristic trimodal DNA methylation distribution, with low or high methylation in individuals homozygotic for the meQTL and methylation levels in between the homozygotic levels in heterozygotic individuals. Although most of these meQTLs are located within several kilobases from the associated CpG site (cis-acting), also trans-acting meQTLs have been identified [115]. meQTLs were found to be

enriched in regulatory intergenic regions (TFBS, enhancers, DNase hypersensitivity regions), while depleted in CGI and promoters [114]. Interestingly, a lot of these meQTLs have been associated with disease. Most of the SNPs found to be associated with human diseases and traits in GWAS are located in intergenic regions, and therefore don't directly affect the protein. It is therefore being hypothesized that disease-associated SNPs may impact gene expression through changes in DNA methylation. Indeed, multiple studies have demonstrated a strong overlap between meQTLs and SNPs which influence expression levels of an associated gene (eQTLs) [116, 117]. Furthermore, meQTLs were found to be enriched for known disease-associated GWAS loci, including schizophrenia [118], bipolar disorder [119], metabolic traits [120] and cancer [121]. In this way, integrating GWAS results with meQTL and eQTL data have been shown to help in interpreting GWAS results and detecting the causal genes and underlying mechanism of complex traits [122-125]. The causal relationship between the SNP, DNA methylation and gene expression is not completely understood. Some studies support the classical model, in which a SNP affects DNA methylation which in turn affects gene expression (DNA methylation as a mediator) [126]. However, other studies also showed the existence of a passive role for DNA methylation, where the SNP influences gene expression and in turn affects DNA methylation [127]. How a SNP may affect DNA methylation is also not completely clear, but is probably due to the disruption of TFBS, thereby leading to a change in the DNA-binding affinity of the TF and subsequently to a change in DNA methylation [128]. TF binding has been shown to have important roles in shaping the DNA methylome [129], and the causal relationship can operate in both directions: either TF binding influencing DNA methylation [130] or else DNA methylation influencing TF binding [131].

Environmental factors. Although twin studies revealed a strong genetic component in explaining DNA methylation variability, the largest fraction of variance (>75%) is explained by unique environmental and/or stochastic factors [103, 108, 111]. Older monozygotic twin pairs showed higher discordance in DNA methylation as compared to younger twins [132-134], and this was shown to be mainly due to unique environmental and/or stochastic factors [134-136]. This apparently random increase in DNA methylation variation with age is called "epigenetic drift" [133, 137]. Interestingly, a recent study seems to confirm this model, showing that the effect of stochastic and environmental influences increases with age [111]. CpG sites which are influenced by environmental factors were found to be depleted in CGI, DHS and proximal promoters, while enriched in CGI shores, shelves, gene bodies and distal promoters [111]. In contrast to this stochastic "epigenetic drift" model, a tissue-independent epigenetic clock (or age estimator) could be constructed based on the methylation levels of 353

CpG sites, indicating the involvement of age-specific regulatory mechanisms (see section of biomarkers for more information) [138, 139].

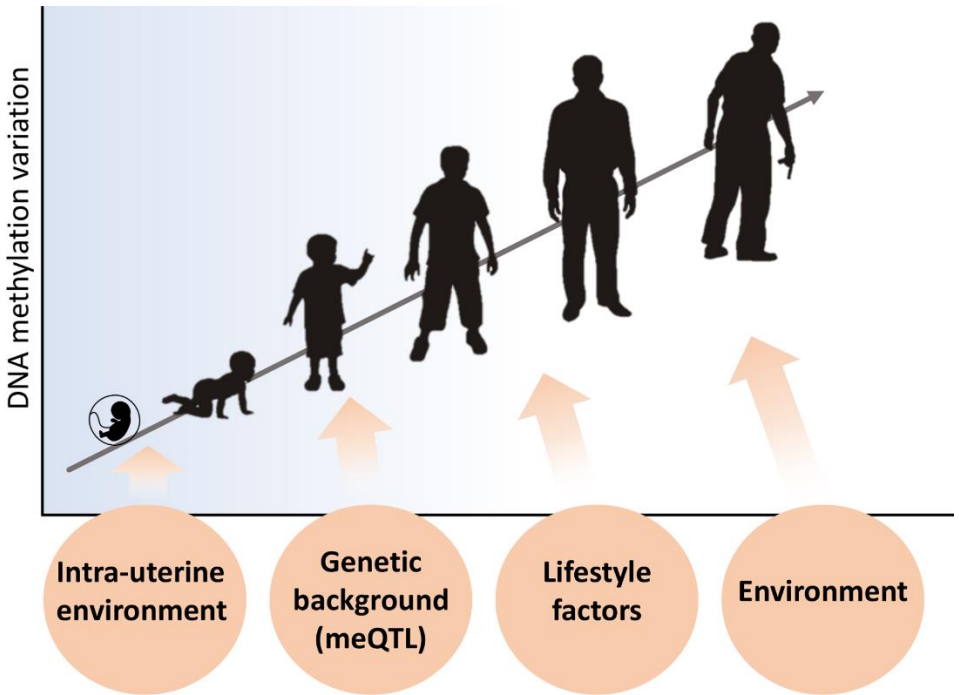


Figure 4: Genetic and environmental influences on DNA methylation. Variability in DNA methylation is influenced by pre- and postnatal environmental factors and genetic factors leading to increased variability with ageing. This figure was inspired from a figure in [154].

Despite the stochastic nature of DNA methylation changes induced by environmental studies, multiple epigenome-wide association studies (EWAS) also found consistent DNA methylation changes associated with specific environmental exposures. Most of these studies are however observational in nature and therefore causality cannot be inferred and confounding not excluded. Multiple environmental and lifestyle factors have been linked with DNA methylation changes, including diet [140], traumatic stress [141], air pollution [142], and alcohol use [143]. One of the most well studied environmental exposures inducing DNA methylation changes is tobacco smoking. In blood, the most frequently reported DNA methylation changes have been reported in the genes AHRR (aryl-hydrocarbon receptor repressor), F2RL3 (coagulation factor II receptor-like 3) and GPR15, which all show a decrease in methylation in active smokers compared to never-smokers [144]. How smoking may induce DNA methylation changes in these genes is not understood, but it seems that some of these sites are also associated with CVD risk, lung cancer and all-cause mortality, indicating their usefulness as potential disease risk predictor [145-148]. Remarkable, quitting smoking may

partially reverse DNA methylation levels towards these of never-smokers depending on the time since quitting [149-152]. Recently, a large fraction of these blood smoking associated DNA methylation changes (including AHRR and F2RL3) could also be replicated in adipose tissue, which is a more informative tissue for metabolic diseases [153]. These studies highlight the impact of environmental exposures on DNA methylation patterns, and also indicate that removing a specific environmental exposure may (after some time) partially reverse aberrant methylation levels towards normal levels. This also points to the potential use of DNA methylation as an environmental biomarker recording past and current environmental exposures. However, whether these findings for tobacco smoking extend to other environmental and lifestyle factors (and could therefore be generalized) should be further confirmed.

Intra-uterine effects and the Developmental origins of Health and Disease

A highly vulnerable period to epigenetic and DNA methylation changes due to environmental conditions is the prenatal life, because of the dramatic dynamic DNA methylation phases occurring during development together with the high cell division and DNA synthesis rate. Multiple studies indeed showed that intra-uterine environmental conditions (often in interaction with genetic factors) shape the neonatal DNA methylome [108, 155-157]. More particularly, it was found that maternal environmental conditions/lifestyle conditions, including smoking, nutrition, pollution, stress and depression have impact on (or at least was associated with) the offspring's DNA methylation profile [158, 159]. Furthermore, because DNA methylation patterns are inherited every cell division, aberrant DNA methylation changes due to intra-uterine environmental conditions may in theory persist until adulthood and impact health later in life [160, 161]. Therefore, DNA methylation is believed to be a potential mediator of the DOHaD paradigm (**Figure 5**). This paradigm proposes that the prenatal and early developmental stages are important contributors of an individual health status throughout its life. This theory was first proposed by Barker in 1990 (also known as Barker's hypothesis) [162], who observed a strong relationship between CVD mortality rates and past infant mortality mainly due to a low birth weight in England and Wales [163]. Low birth weight was here used as a surrogate for fetal undernutrition. This observation (which was later replicated in other geographic regions) formed the basis of the developmental origin hypothesis [5]. Based on a large set of observations this hypothesis is now widely accepted and also extends towards other chronic disease, including diabetes, hypertension, obesity and even mental diseases [164, 165]. It has been hypothesized that intrauterine environmental conditions program the fetus to prepare and adapt the offspring for a similar exposure postnatally. However, if there is a difference in pre- and postnatal life, as is the case

during an acute famine, this mismatched adaptation may evolve towards disease development later in life [166].

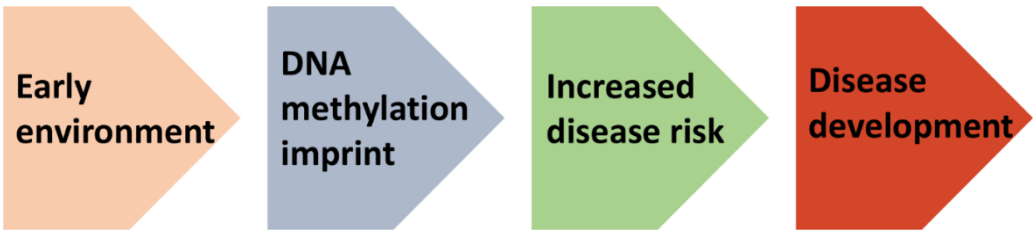


Figure 5: DNA methylation as a mediator of DOHaD. In the classical view, pre- and early postnatal life environmental factors induce DNA methylation changes leading to increased disease risk and disease development in adult life.

Due to their cellular memory function, DNA methylation and other epigenetic mechanisms seem to be promising mediators for translating past in-utero aberrant environmental conditions in diseases later in life. Although in humans there is a wealth of studies associating prenatal environmental conditions with DNA methylation changes in postnatal life, it is very difficult to causally link this with disease and health in later life. A well-known human study is the Dutch famine cohort study, which studies the effect of undernutrition during gestation on adult health [167, 168]. The Dutch famine was a consequence of a food transport embargo by the Germans during the end of the WO II (winter of 1944-1945). Maternal undernutrition resulted in a reduction of glucose tolerance in the offspring independent of the exposure time (early, mid, or late gestation). In contrast, only adults exposed to famine during early gestation had a more atherogenic lipid profile, and higher risk for obesity and coronary heart disease. Therefore, early gestation seems to be the most vulnerable period.

There is strong evidence now that DNA methylation is important in linking prenatal famine and disease status in later life. Early target-genes studies found a reduction in blood DNA methylation of the imprinted IGF2 gene [169], and other growth and metabolic disease associated genes, including LEP, INSIGF, IL10, ABCA1, GNASAS and MEG3 [170]. This was measured in whole blood samples of adults (~60 years of age). More recent genome-wide studies replicated these results and found differentially methylation especially in genes involved in growth and metabolism-related pathways [171]. Although, these results suggest that DNA methylation may be an important factor in linking adverse prenatal environmental conditions (in this case malnutrition) and disease risk in later life, it is very difficult to elucidate causality from this observational study. For example, DNA methylation could be affected by the famine, but is maybe not the cause of the phenotype seen in adults, which is affected by another mechanism. Another possibility is that the disease phenotype itself results to

the observed DNA methylation changes. A recent study however, supports the classical view by showing that DNA methylation is a mediator between prenatal famine exposure and adult BMI and triglyceride levels [172]. However, to completely rule out reverse causation, longitudinal study designs are needed. What also became clear from these studies is that the timing of the prenatal exposure is crucially important. For example, these DNA methylation changes in adults were only seen when they were exposed early in gestation (week 1-10), and not in mid- or late gestation [170, 173]. This implicates that early gestation period may be a critical time-window during which prenatal environmental conditions may impact the human methylome.

Inter- and transgenerational DNA methylation inheritance

The methylation changes seen in the offspring due to prenatal environmental exposures is an example of intergenerational DNA methylation inheritance. In addition, evidence (mainly in animal models) is also available that environmental effects of the father may also impact offspring's health [174]. In humans, an often cited study in this context, is the Överkalix study, which uses historical records of harvests and food prices in Överkalix, an isolated community in northern Sweden, to link food availability with health outcome in child's and grandchild's offspring. They found that CVD mortality was low when food was not readily available during the father's slow growth period (period before puberty) [175, 176]. Interestingly, they could also find a transgenerational effect: if the paternal grandfather was exposed to an abundance of food during his slow growth period, diabetes mortality increases [175, 176]. In addition, this effect was sex-specific as it was only observed in the male grandchildren [177]. A recent large-scale study using 9,039 grandparents, 7,280 children and 11,561 grandchildren from the Uppsala Multigeneration Study, confirmed this overall result in that paternal grandfather's food access predicts his male, but not female, grandchildren's all-cause mortality [178]. Also other environmental cues than nutrition may impact the son's health outcome exemplified by the Avon Longitudinal Study of Parents and Children which found that fathers who smoked during their prepubertal period have sons with an higher BMI [174, 179].

These studies indicate that beside maternal lifestyle factors during pregnancy also paternal lifestyle conditions may impact the offspring's health outcome. Whether DNA methylation is a possible mediator for this intergenerational inheritance is not completely understood. In contrast, to the DOHaD paradigm where environmental exposures can directly influence the fetus, paternal and preconceptional exposures should be transmitted through the germ line. This means that environmental factors should induce DNA methylation changes in the germ cells, and should escape from

erasure during the global reprogramming phase after fertilization. For example, imprinted centers, which are crucial in establishing parent-of-origin-specific gene expression in certain genes, are not erased during embryogenesis (**Figure 3**). Interestingly, obesity was shown to influence methylation pattern of imprinted genes in sperm cells [180]. In addition, in another cohort parental preconceptional obesity was found to alter methylation status of the same imprinted genes in the offspring [181], indicating a possible intergenerational effect. Beside imprinted genes, also other genomic regions, especially repeat sequences including SINEs, VNTRs, ALUs and tandem repeats, were found to be protected from this global demethylation phase [182]. Another study showed that Holocaust survivors had a different FKBP5 (a regulator of glucocorticoid receptor sensitivity) methylation status in blood as compared to control subjects, and the same CpG site was also found to be differentially methylated in the offspring [183]. In general however, most studies on inter- and transgenerational epigenetic inheritance is done in animal models and human evidence is rather limited. Furthermore, in humans it is very difficult to distinguish between real epigenetic inheritance and other confounding interactors.

The grandparent's environmental effect on the grandchildren seen in for example the Överkalix study, is an example of transgenerational inheritance, which is different from intergenerational inheritance by that the effect of the exposure is transmitted through multiple generations and affects individuals who were never exposed to the environmental cue (**Figure 6**). Therefore, the effect of the exposure should be at least detectable in the F3 generation, as environmental factors in pregnant mothers (F0) already influence the F2 generation through the germ cells of the F1 generation. In non-pregnant mothers or fathers (F0), the effect of the exposure should be observable in at least the F2 generation. On the assumption that DNA methylation is a driver for transgenerational inheritance, this also implies that DNA methylation changes induced by prior environmental conditions should survive two global demethylation phases: both during embryogenesis after fertilization and during the formation of gametes. Imprinted genes for example retain their methylation state during embryogenesis but are removed during gametogenesis [96]. However, just like during the global demethylation phase after fertilization, some repeat elements are resistant to global demethylation, especially evolutional young and currently active retrotransposons including SVAs and L1 LINE repeats [90]. More interestingly, the same study also found repeat-poor regions that escape genome-wide DNA demethylation in primordial germ cells. These regions were mainly located in enhancers, CGI, promoters and gene bodies and enriched in genes associated with obesity-related traits, schizophrenia and multiple sclerosis. This is interesting, because it implies that, at least theoretically, transgenerational inheritance through a DNA methylation mechanism is possible.

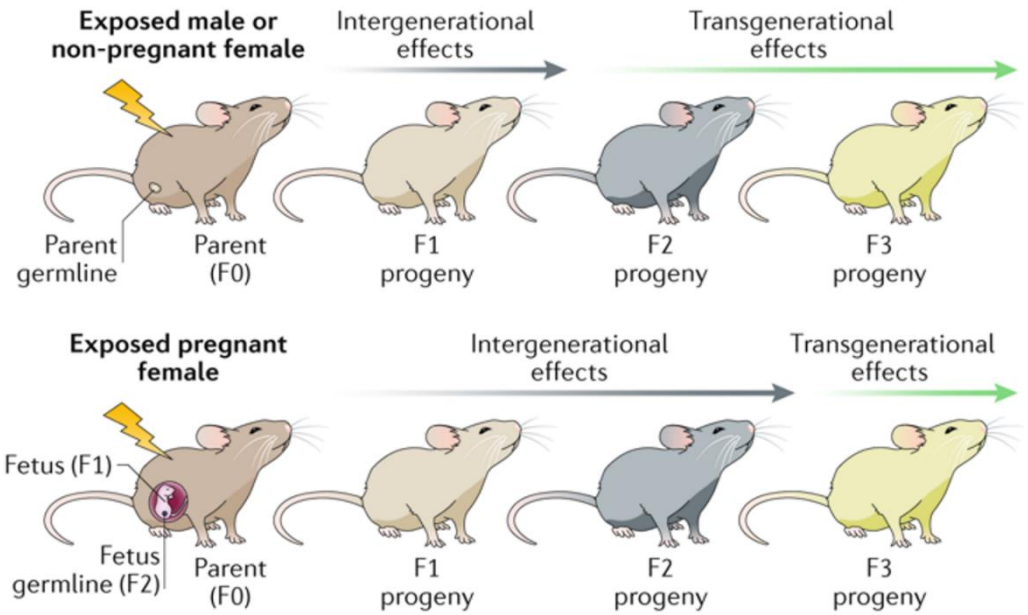


Figure 6: Inter- and transgenerational epigenetic inheritance. In males and non-pregnant females, DNA methylation changes should be at least visible in F2 to be transgenerational. In pregnant females, transgenerational DNA methylation changes should be at least detectable in the F3 generation. Figure was adapted from [184].

In humans, DNA methylation changes that persist through multiple generations have not yet been identified, however, in other mammals, like rats and mice some evidence exists. One of the earliest studies demonstrating transgenerational inheritance was found in rats exposed to the endocrine disruptor vinclozolin [185]. Exposure during gestation resulted in an increased incidence of infertility in the F1 males. Furthermore, this phenotype was transmitted through the male germ line up to the F4 generation, and correlated with DNA methylation changes [185]. However, the role of these DNA methylation changes to the development of infertility in these rats remains uncertain. A more recent study, for example, showed that the sperm DNA methylation changes in the F1 generation was different from the DNA methylation changes in the sperm of F3 generation [186]. In support of this finding, a study found no persistent DNA methylation changes in the F2 and F3 generation, while the F1 germline did show methylation changes due to gestational vinclozolin exposure indicating that the methylation changes induced in the F1 generation are corrected in the germline during the global epigenome reprogramming [187]. Although here we focused on DNA methylation, evidence is emerging that also other epigenetic factors, including histone modifications and small RNAs may contribute to inter- and transgenerational

inheritance and are probably working together to mediate epigenetic transgenerational inheritance [184, 188, 189].

Linking cellular metabolism to epigenetic profiles

Although a lot of studies could associate environmental exposures (both *in utero* as during the lifespan) with specific DNA methylation changes, how the environment influences epigenetic profiles is not completely understood and is less well investigated. Cellular metabolism may be an important player in this regard because most of the epigenetic machinery depends on the availability of specific metabolites [190-192]. Therefore changes in cellular metabolic states due to environmental factors may change intracellular metabolite and cofactor levels influencing enzymatic activity of chromatin modifiers which subsequently change epigenetic patterns. For example, histone acetylases (HATs) depend on acetyl-CoA as a substrate for the acetylation reaction. Cellular levels of acetyl-CoA are strictly dependent on energy status: when energy production is high, acetyl-CoA is upregulated, in contrast, when cellular metabolism is slow, acetyl-CoA levels are reduced. The sirtuin family of histone deacetylases (SIRT1-7) is dependent on NAD+, an important electron carrier during redox reactions. LSD histone demethylases use FAD as cofactor, which is produced from riboflavin (vitamin B2).

DNA (and histone) methyltransferases rely on the universal methyl donor SAM. The reaction between methionine and ATP produces SAM. A methyltransferase reaction donates the methyl group of SAM to an acceptor macromolecule (for example cytosine or an histone) yielding S-adenosylhomocysteine (SAH), which is hydrolyzed to homocysteine and adenosine. Homocysteine can subsequently close the cycle by converting back to methionine via the transfer of a methyl group from 5-methyltetrahydrofolate (5-MTHF). Therefore the formation of SAM is dependent on folate and other methyl donors which must be obtained from the diet (**Figure 7**) [140]. Dietary methyl donors include methionine, folate, betaine, choline, vitamin B2, B6 and B12. A large number of rodent studies have shown that dietary methyl donor intake may influence DNA methylation, providing a link between nutrition and DNA methylation [140]. Mice carrying the viable yellow allele of the agouti (Avy) gene have been a valuable model in this regard. This allele harbors an insertion of an intracisternal A-particle (IAP) upstream of the agouti gene. The level of DNA methylation of this IAP determines the expression level of the agouti gene, and is stochastically established early in development and is therefore called a metastable epiallele. Hypomethylation of the inserted IAP result in an active allele which gives rise to a yellow fur and a phenotype of obesity, type II diabetes and predisposition to tumors. In the wild-type

condition, the allele is methylated and silenced resulting in lean dark-brown colored mice (called pseudoagouti). Interestingly, the methylation state of this IAP can be influenced by maternal methyl donor supplementation, shifting towards pseudoagouti offspring [193, 194]. A similar result was found in another metastable epiallele, the Axin Fused (AxinFu) locus [195], which shows variability in tail morphology. These results suggest that methyl donor levels may influence DNA methylation states. Similar results are now also found in humans where paternal and maternal methyl-group donor intake before and during pregnancy influences DNA methylation in the offspring [196-198]. In contrast, a large cross-sectional study (n>5,000) didn't find any association between methyl-donor intake and DNA methylation in blood [199].

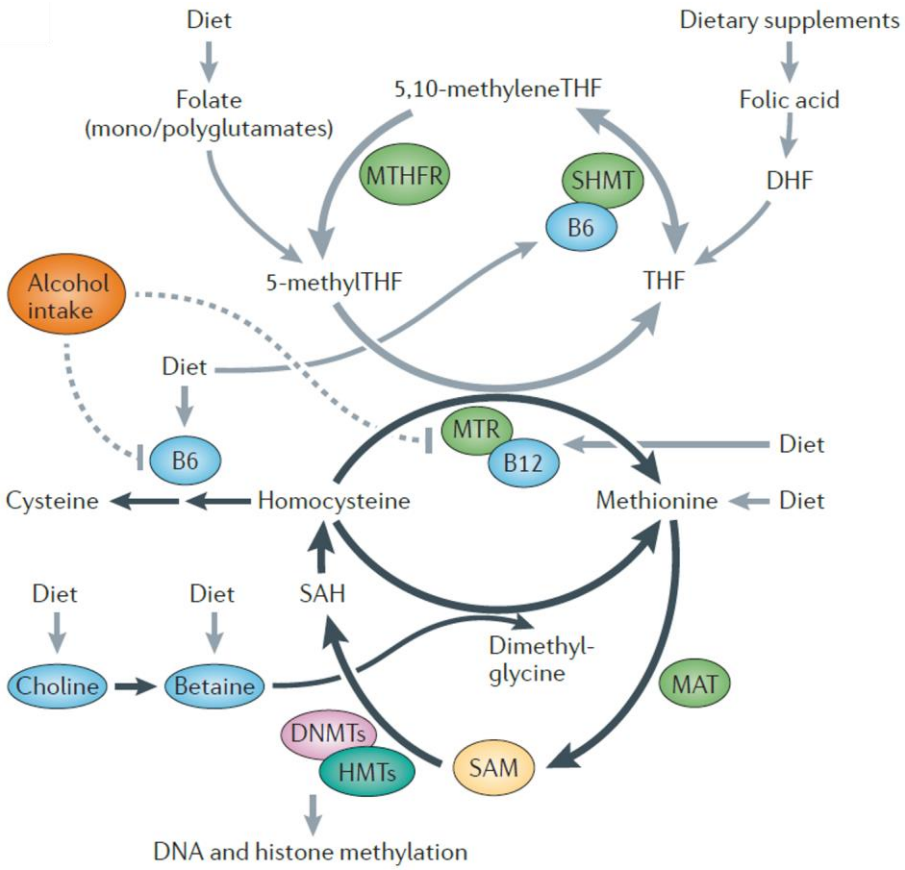


Figure 7: Dietary impact on SAM levels and DNA methylation. SAM is generated by the methionine (or one-carbon) cycle (thick black arrows). Methyl groups are retrieved from dietary sources, including folate, choline and betaine. Vitamin B6 and B12 are two important cofactors involved in SAM biosynthesis. Figure adapted from [9].

The TET enzymes, which are important in active DNA demethylation, and the Jumonji-domain containing histone demethylases, are 2-oxoglutarate-dependent dioxygenases. 2-Oxoglutarate is an intermediate of the TCA cycle and is produced from isocitrate by

isocitrate dehydrogenase 2 and 3 (IDH2 and -3) in the mitochondria. IDH1 catalyzes the formation of 2-oxoglutarate in the cytoplasm. IDH mutations are frequently observed in several cancers, including gliomas and hematological malignancies [200]. Mutated IDH enzymes lead to the formation of 2-hydroxyglutarate (2-HG), which is a competitive inhibitor of 2-oxoglutarate, and therefore alters the activity of TET enzymes [201]. As expected, this diminished TET activity results in a DNA hypermethylation phenotype in this tumors [202]. In addition, mutations in succinate dehydrogenase (SDH) and fumarate dehydrogenase (FH) which results in the accumulation of the TCA cycle intermediates succinate and fumarate respectively, can also inhibit TET enzyme activity and consequently leads to genome-wide DNA methylation changes [203]. Interestingly, ascorbate (Vitamin C), an essential micronutrient, is a cofactor for TET and has been shown to impact the DNA methylome [204]. TET enzymes are also dependent on oxygen and iron Fe(II). For example, tumor hypoxia has been shown to reduce TET activity due to a decreased oxygen availability and causes promotor hypermethylation [205]. However, in another study hypoxia induces TET1-mediated global hydroxymethylation [206]. DNA methylation is also responsive to oxidative stress as ROS levels reduce TET enzyme activity [207] and lead to a global decrease in DNA hydroxymethylation [208].

Beside changing cellular metabolic and oxidative state, some nutritional compounds also has been shown to directly interfere with the enzymatic activity of various epigenetic modifiers [209]. As an example, (-)-epigallocatechin-3-gallate (EGCG), the major polyphenol from green tea, can inhibit DNMT activity through the formation of hydrogen bounds in the catalytic pocket of DNMT resulting in DNA methylation changes in cancer cell lines [210]. Other phytochemicals that may impact epigenetic enzymes include, among others curcumin, epicatechin, genistein and resveratrol. However, we should keep in mind that most of these studies were performed *in vitro*, and the *in vivo* efficacy has not been proven yet.

The above mentioned studies clearly show that cellular metabolic and redox state is important in maintaining DNA methylome profiles, and that alterations in the levels of certain metabolites may impact the enzymatic activity of the DNA methylation machinery leading to global methylation changes. However, environmental EWAS also identified multiple specific CpG sites or small genomic regions which are consistently differentially methylated among the exposed sample group. Why are certain CpG sites more vulnerable to DNA methylation changes, and are other rather resistant? An answer to this question lies probably in the high correlation seen between TF binding and DNA methylation patterns. In the classical view, TFs are mainly bound to unmethylated binding sites, while DNA methylation of that site may affect TF binding [211]. Although the causal relationship between TF binding and DNA methylation is not

completely understood and works probably in both directions [212], environmental exposures which influence TF binding are predicted to also result in locus-specific DNA methylation changes [213]. Therefore environmental-induced epigenetic changes may also be due to the inhibition or activation of certain TFs.
This complex link between TFs and DNA methylation will be further explained in more detail in the next section when we explain its main role as maintainer of gene expression profiles. In addition, we will mention the involvement of DNA methylation in other alternative functions.

Functions of DNA methylation

Based on studies in the 70'-80', a dogma emerges where DNA methylation is associated with transcriptional repression. DNA methylation has indeed been shown to directly silence gene expression, however, studies from the last two decades demonstrated a much more complex role of DNA methylation which seems to be dependent on the genomic context. Here, I will describe the mechanisms how DNA methylation maintains gene expression profiles, controls genomic imprinting, alternative splicing and genomic stability.

The complex relationship between DNA methylation and gene expression

The high association between CGIs and promoter regions indicates the importance of DNA methylation in transcriptional regulation and initiation. Indeed, DNA methylation was first described as a transcriptional silencer mark. Now, we know that the relation between DNA methylation and gene expression is much more complex and that both negative and positive correlations with gene expression exist (**Figure 8A**) [214].
Approximately half of the CGIs are located at gene promoters and are most often unmethylated, even when the corresponding gene is not expressed [215]. Most of the inactive CGI promoters are therefore not methylated but become trimethylated at lysine 27 of histone 3 (H3K27me3) mediated by Polycomb proteins [216]. Active CGI promoters on the other hand are associated with the active H3K4me3 histone mark [216]. However, examples do exist of methylated CGIs leading to stable transcriptional silencing of the corresponding gene, and it was estimated that in human brain approximately 3% of promoter CGIs are methylated [87]. This is especially the case in germline-specific genes which become *de novo* methylated and repressed during differentiation from germ cells into somatic cells [215, 217]. Also in X-chromosome inactivation and imprinting, DNA methylation of CGI leads to the repression of gene

expression. Thus, we can say that methylation at CGI promoters indeed leads to silencing of the corresponding gene, but in somatic cells this is a rather marginal phenomenon occurring mainly in germ-line specific genes and during X-chromosome inactivation, and that other mechanisms (i.e. histone modifications) are needed to inactivate the gene.

The other half of the CGIs are located intergenic and intragenic and are much often methylated, and much more variable across different tissues [88]. Interestingly, most of these orphan CGIs share characteristics of functional promoters, including transcriptional initiation during development [218]. Indeed, intragenic DNA methylation may play a role in regulating alternative promoter usage in a cell-specific way (**Figure 8B**) [87]. For example, an intronic CGI was important in regulating cell-type specific gene expression of the MCJ gene, and not the promoter CGI [219]. Thus, variable intragenic DNA methylation across cell types may regulate cell-type specific transcription [220], and high intragenic DNA methylation prevents aberrant transcription initiation in the gene body [221]. Intragenic DNA methylation is established by DNMT3B which is recruited to the gene body by the H3K36me3 histone mark [221]. Another mechanism was recently proposed, where intragenic CGI methylation is dependent on transcription through this CGI [222].

Promoters with lower CpG density are much more frequently methylated [215], but not always result in the silencing of the corresponding gene [223]. Interestingly, a recent study discovered that distal upstream CGIs are better correlated with gene expression in non-CGI promoters, and that these CGIs may function as alternative promoters in active genes with a methylated promoter [223]. Although the correlation between DNA methylation and gene expression non-CGI promoters is much more complex than in CGI promoters, there is also evidence that DNA methylation at these promoters may also directly silence gene expression, as demonstrated for the RUNX3 promoter [224].

What is the exact role of DNA methylation in gene expression silencing? As we have previously seen, *de novo* DNA methylation during development is almost always preceded by histone methylation, suggesting that DNA methylation is rather a secondary mechanism in gene silencing. For example, the histone methyltransferase G9a ensures the silencing of the pluripotency gene OCT4 by H3K9 methylation. In addition, G9a can recruit DNMTs which leads to *de novo* methylation and long-term silencing preventing embryonic reprograming [76]. In another study, it was found that the Polycomb group protein EZH2, which sets the repressive H3K27me3 marks, interacts with DNMTs, leading to DNA methylation [225]. In addition, *de novo* methylation of promoters in embryonic stem cells mainly harbor the H3K27me3 histone mark [226]. Another example where DNA methylation has rather a passive role in gene silencing is during X-chromosome inactivation where gene silencing precedes

DNA methylation [227]. These examples indicate that DNA methylation is often a secondary mechanism in transcriptional silencing and ensures stable gene repression thereby maintaining cellular identity and preventing cellular reprogramming. Recent studies, however, also demonstrated an active role of DNA methylation in establishing chromatin states. For example, reduction of DNA methylation levels leads to a decrease of the H3K27ac active histone marks at enhancers [228]. In agreement, global demethylation in mouse embryonic stem cells changes H3K27ac, H3K27me3 and H3K4me1 occupancy in enhancers and promoters, while there was no impact on H3K4me3 [229]. Interestingly, these changes in histone mark deposition were re-established upon the addition of DNMTs.

Readers of DNA methylation

One way how DNA methylation is translated in the stable repression of a gene, is by the binding of proteins with a MBD [230]. The MBD protein family consists of five members (MeCP2, MBD1, MBD2, MBD3 and MBD4) which bind methylated DNA in a sequence independent way (except for MBD3 which does not bind methylated DNA). Binding of MBD proteins on methylated DNA leads to the recruitment of histone deacetylases and methylases which subsequently alter chromatin structure and finally repression of gene expression (**Figure 8A**). MBD-containing proteins can therefore be seen as readers of DNA methylation.

Besides the well-known MBD proteins, evidence is emerging that also certain TFs can recognize and bind methylated DNA (**Figure 8A**). MBD proteins mainly bind high dense methylated regions [231], and in the traditional view, methylation disrupts the binding of other TFs which need unmethylated DNA to bind [211]. Indeed, there is a strong correlation between TF binding and DNA methylation which also indicate the importance of DNA methylation in more distal regulatory regions, like enhancers in regulating gene expression [129]. Enhancers and TFBS are indeed more variable methylated across cell types [86], and often have low/intermediate methylation levels (~30%) indicating dynamic regulation [129]. Although this DNA-TF binding disruption mechanism plays an important role in DNA methylation-dependent gene expression, recent studies demonstrated the involvement of other mechanisms, where certain TF can bind methylated DNA in a sequence-specific way [211]. Examples are CTCF, KLF4, CEBPb and Kaiso, which can bind both methylated and unmethylated sequences. Of note, often the methylated and unmethylated sequence motifs are different. The biological consequences of this new paradigm are not yet understood, but could provide another explanation for the positive correlation observed between promoter methylation and gene expression in certain genes. In support of this hypothesis, a

recent study found that positive correlated regions harbor different sequence motifs than negative correlated regions [214], indicating the involvement of different TFs.

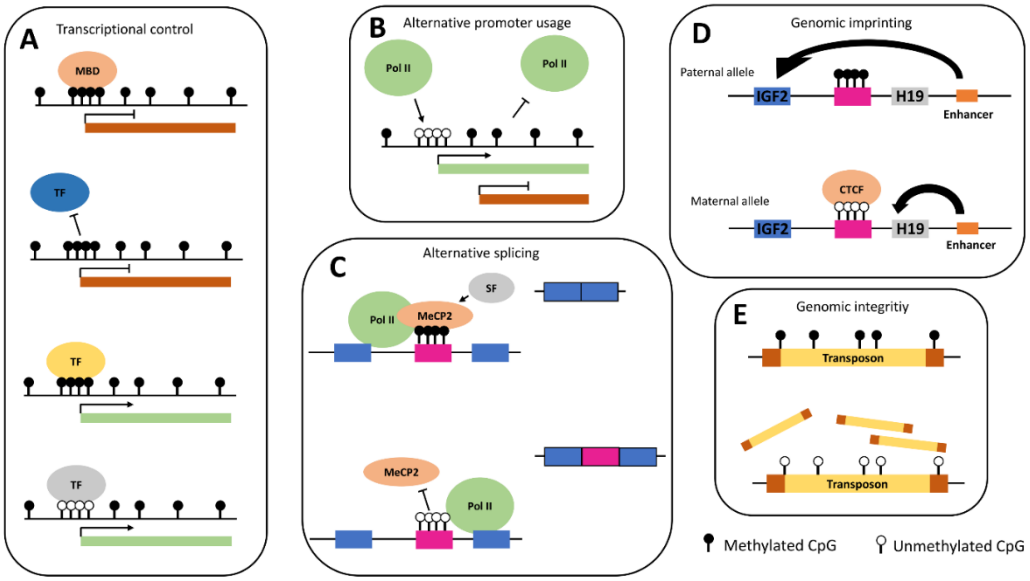


Figure 8: Functions of DNA methylation. A) Transcriptional control: MBD-containing protein binding methylated DNA, TF binding either methylated or unmethylated DNA, or methylation prevents TF from binding. **B)** Alternative promoter usage: intragenic DNA methylation: high gene body methylation prevents spurious transcription initiation. **C)** Alternative splicing: MeCP2 binding on methylated alternative exon causes Pol II pausing and recruitment of splice factors (SF) leading to splicing. **D)** Genomic imprinting: example of IGF2-H19 imprinted region. Methylation of imprinted center prevents the CTCF insulator from binding leading to IGF2 transcription on the paternal allele, while IGF2 is silenced in the maternal allele. **E)** Genomic integrity: DNA methylation prevents transposons from 'jumping'.

TF binding and DNA methylation are thus strongly correlated, but the causal relationship between the two factors are not yet completely clear. Intuitively, DNA methylation affects the interaction of TFs with DNA, which implies an active role of DNA methylation. However, in contrast, more recent research indicates that TF binding can also induce DNA methylation changes. For example, deletion of the TF REST leads to an increase in DNA methylation at its binding sites [130]. On the other hand, genetic deletion of the three DNMTs in mouse ES cells resulted in the creation of novel NRF1 binding sites which were highly methylated in the wild-type cell line, indicating that removal of DNA methylation creates new binding sites [131], which is in agreement with a model where DNA methylation restricts TF binding. In contrast, other TF-DNA bindings are mainly insensitive to the removal of DNA methylation, as recently exemplified for CTCF [232]. Furthermore, it has been shown that the binding sites of methylation-sensitive factors (like NRF1) are kept in a hypomethylated state by the adjacent binding of methylation-insensitive factors (like CTCF and REST) which can

change the local methylation status [131]. Thus, these results suggest a complex model where both methylation-sensitive factors, methylation-insensitive factors and (de)methylation enzymes cooperate in regulating DNA methylation levels, TF binding and subsequent gene expression profiles.

Gene body DNA methylation regulates alternative splicing

Alternative splicing is an evolutionary conserved process in which multiple mRNAs and proteins are formed from a single gene due to the inclusion or exclusion of different exons during and after mRNA formation [233]. In humans, almost all exon-containing genes are alternative spliced which result in an increased proteomic diversity important in cell and tissue type specificity and responses towards external stimuli. Although multiple mechanisms exist, in general, alternative splicing is regulated by RNA-binding splice factors which interacts with specific cis-acting consensus sequences (for examples splice sites at exon-intron boundaries). The strength of the consensus sequences determines the inclusion potential of the exon [234].

A role of intragenic DNA methylation in alternative splicing is emerging (**Figure 8C**) [234]. For example, it has been demonstrated that exons have higher DNA methylation levels than introns [235]. Furthermore, DNMT knock-out embryonic stem cells showed altered alternative splicing in both directions: inclusion and exclusion of alternative exons [236]. A similar result was found when treating cancer cells with the DNMT inhibitor 5-Aza-2'-deoxycytidine (5-Aza-CdR) [237]. A direct causal role of DNA methylation in splicing was demonstrated using the targeted addition or removal of DNA methylation at alternative spliced exons by a CRISPR-Cas9 enzyme fused with a DNMT and/or TET enzyme [238]. DNA methylation plays probably a 'fine-tuning' role in splicing and multiple mechanisms have been proposed. For example, binding of the MBD-containing protein MeCP2 to methylated exons results in the pausing of Pol II and subsequent inclusion of the targeted exon [239]. A similar role was demonstrated for CTCF. CTCF is a methylation-sensitive TF only binding at unmethylated DNA motifs [240]. Just as with MeCP2, CTCF exon binding leads to Pol II pausing and inclusion of the alternative exon, while methylation of this exon prevents CTCF from binding resulting in the exclusion of the exon. Of interest, DNA methylation may therefore enhance exon inclusion (e.g. MeCP2) or prevent exon inclusion (e.g. CTCF), indicating a complex role of DNA methylation in splicing. A different mechanism, involving heterochromatin protein 1 (HP1), was recently demonstrated [236]. Here, DNA methylation induces H3K9me3 histone modifications at nearby histones of alternative exons which serve as docking sites for HP1 leading the recruitment of splice factors to the methylated exon. So, DNA methylation seems to regulate splicing using two

different mechanisms: a 'kinetic' mechanism where Pol II elongation is reduced, and a 'recruitment' mechanism where specific splice factors are recruited. A recent interesting study however, showed that these two models are linked and that MeCP2 is also able to recruit splice factors and regulating intron retention [241].

Genomic imprinting

Genomic imprinting is a remarkable epigenetic phenomenon where only one of the alleles of a gene is active and expressed, while the expression of the other gene copy is suppressed, and this depends on whether the gene was maternally inherited or paternally inherited [102, 242, 243]. In human genomes, approximately 160 genes are imprinted [244]. Most of these genes are involved in the growth of the embryo, placenta and neonate. In addition, imprinting has also a role in the survival of neonates by regulating feeding, maintenance of body temperature, regulation of metabolism and behavior. Remarkable, some imprinted genes are monoallelic expressed in specific cell types [245] or certain developmental windows. For example KCNQ1 and KCNQ1OT1 were found to be monoallelic expressed only in fetal tissues and not in adult tissues [246]. Although several hypotheses have been postulated [247-249], why imprinting exists is not yet completely understood. The most popular theory is the kinship or parental conflict theory which is based on differences in interest between paternal and maternal genes [250, 251]. The paternal genome would benefit from maximal acquiring maternal resources during the growth of a fetus. This will of course benefit the offspring but comes with a higher cost to the mother. Therefore, the maternal genome would more benefit from a thrifty way of using maternal resources during embryonic growth, because in this way the chance of bearing future offspring will be higher. This is also reflected in the function of genes which are paternally and maternally expressed. Many paternally expressed imprinted genes enhance embryonic growth while many maternally expressed imprinted genes repress embryonic growth. In addition, imprinting is only seen in placental mammals and marsupials and not in egg-laying mammals [252].

DNA methylation differences between the two alleles of an imprinted gene are crucial in establishing and maintaining this parent-of-origin-dependent gene expression. Typically, imprinted genes are clustered in genomic regions which can vary in size from < 100 kb to several megabases and contain 2-15 genes. Each cluster contains an imprinted control region (ICR) which is differentially methylated between the maternal and paternal allele in the germ line and are therefore called germline DMRs (gDMRs). During development, this gDMRs can result in the formation of additional DMRs which are called secondary or somatic DMRs. At the beginning of gamete formation, the old

imprints are first removed during the genome-wide demethylation phase and are subsequently replaced by new imprints. How these specific gametic DMRs are recognized for methylation is less clear, but may involve transcription and histone modifications [253, 254]. After fertilization, imprinted DMRs are protected from genome-wide DNA demethylation, by trans-acting factors including ZFP57-KAP1 [255, 256], PGC7 [257] and G9a/GLP [258] which may recruit histone modifiers and DNMTs and/or prevent TET enzymes from binding.

In almost all clusters, at least one lncRNA gene is involved and is often crucial in regulating gene expression of the imprinted genes in the cluster. In some imprinted clusters, the lncRNA gene expression is regulated by methylation in the ICR. Expression of the lncRNA can subsequently silence the protein-coding genes in the cluster in cis. KCNQ1 is an example. ICRs can also function as insulators by binding CTCF and subsequently separating the enhancer from the protein-coding genes. Methylation of the ICR, however, prevents CTCF from binding leading to the expression of the protein-coding genes. An example is the IGF2-H19 cluster (**Figure 8D**).

The importance of imprinting is further implied by a number of imprinted disorders, including Prader-Willi syndrome, Angelman syndrome, Beckwith-Wiedemann syndrome, pseudohypoparathyroidism types 1a and 1b, and Silver-Russel syndrome [259]. These syndromes are mainly the result of loss of imprinting by genetic and epigenetic alterations of specific imprinted genes/regions. For example, the majority of Silver-Russel syndrome patients have an hypomethylation of H19 DMR which results in the silencing of IGF2 and biallelic expression of H19. Another part of the patients, however, show maternal uniparental disomy (UPD) for chromosome 7 in which both copies of the complete or part of the chromosome is inherited from the mother. A lot of these disorders leads to developmental abnormalities, such as growth retardation (Silver-Russel, Prader-Willi) or overgrowth (Beckwith-Wiedemann). Imprinting is also involved in common diseases, including obesity, IUGR, and cancer [243].

Maintaining genomic integrity

Repression of repeats. Most of the human genome consists of repeat elements of which transposons are a major subgroup and represents about 50% of the genome. Transposons are genomic repeat elements which can “jump” from one location to another and may therefore cause diseases when it “lands” in a gene or regulatory region. However, transposition may also aid in genome evolution. Retrotransposons (like LINEs, SINEs and LTR) need transcription to RNA to insert into a new chromosomal sites. DNA transposons, on the other hand, “jump” by a cut-and-paste mechanism

without RNA transcription. In humans, only a small part of the retrotransposons are active and able to transpose.

Transposons are highly methylated. DNA methylation plays an important role in silencing of transposons and preventing it from "jumping" (**Figure 8E**). For example, the mouse intracisternal A particle (IAP) LTR retrotransposons are activated in DNMT1 deficient mouse embryos [260]. Preventing *de novo* methylation by deleting DNMTs results in massive transcription of transposons in male germ cells. Especially deletion of DNMT3L was found to have a dramatic effect [97, 261]. Recently, a new *de novo* DNA methyltransferase DNMT3C which arose from a duplication of DNMT3B was discovered and seems to be much more important in silencing retrotransposons than the other *de novo* DNMTs [262]. piRNAs are crucial in guiding the DNA methylation machinery to transposons for *de novo* methylation, at least in LINEs [101, 263-265]. piRNAs are small silencing RNAs which interact and are preprocessed by PIWI proteins and can bind antisense to transposon transcripts leading to the recruitment of DNA methylation enzymes and histone modifiers [266]. In addition, piRNAs can also silence transposons post-transcriptional by a slicing mechanism [266]. In contrast to LINEs, SINEs are not methylated and silenced by a piRNA-dependent pathway. Although also the majority of SINEs are highly methylated [267], loss of DNA methylation has little impact on SINEs expression, indicating that DNA methylation may not be a primary mechanism in SINE silencing [268]. On the other hand, inhibition of histone methyltransferases does promote SINE expression [268].

***Genome stability*.** Global hypomethylation is seen in a lot of cancers and has been linked with an increase in genomic instability and mutation rate [269-272]. Another example where the role of DNA methylation in maintaining genome integrity is clearly demonstrated, is in patients with the ICF (Immunodeficiency, Centromere instability and Facial anomalies) syndrome. This rare recessive autosomal genetic disease is caused by mutations in the DNMT3B gene and shows large hypomethylated regions in satellite DNA and pericetromeric repeats which are normally highly methylated. This hypomethylation leads to chromosomal breakage and DNA rearrangements of chromosomes 1, 9 and 16 [51, 273]. In another study, loss of DNMT resulted in decreased DNA methylation levels in subtelomeric regions leading to elongated telomeres and increased telomeric recombination [274]. These examples clearly show that DNA methylation not only is important in maintaining transcriptional programs, but also safeguards genomic integrity.

The DNA damage repair system is crucial for the cell to prevent the accumulation of mutations and development of tumors. Multiple studies have shown that DNMT1 is an important part of the DNA damage repair response, in a process which is independent

of its DNA methyltransferase enzyme activity [275]. DNMT1 was shown to be recruited to double strand DNA breaks [276], interacts with repair proteins and control DNA damage responses [277]. In addition, DNMT1 also has a role in mismatch repair, which is a system that detects and repairs base-base mismatches and thereby preventing microsatellite instability. During replication of microsatellite repeats, DNA polymerases may “slip” and subsequently lead to an expansion or contraction of these repeats. DNMT1 deficiency in mouse embryonic stem cells was shown to lead to a higher DNA slippage rate at mononucleotide repeats resulting in microsatellite instability [278, 279]. Again, this process was found to be DNA methylation-independent [280]. The higher mutation rates seen with DNMT1-deleted embryonic mouse stem cells [271], may therefore also be the consequence of the lack of DNMT1 and its role in DNA repair instead of the global DNA hypomethylation.
Recent evidence also demonstrates a link between TET enzymes, hydroxymethylation and DNA repair. 5hmC is accumulated at sites of DNA damage [281]. In addition, different TET knock outs lead to impaired DNA damage response [282], increased chromosome segregation defects during mitosis [281] and more DNA strand breaks [283].

DNA methylation in disease

In analogy to genetic mutations, epi-mutations also exist which are associated with specific diseases. For years, cancer epigenetics dominated the field and it is only recently that adverse epigenetic alterations also have been described in other complex diseases. Detection of disease-associated DNA methylation changes may increase our understanding of mechanisms underlying complex diseases. Furthermore, DNA methylation patterns may be used as biomarkers and may reveal interesting therapeutic targets.

DNA methylation in cancer

Initially, oncology was mainly focusing on genetic alterations, but now it is clear that also epigenetic changes are crucial both in initiation and progression of cancers. There exists a remarkable interplay between genetic and epigenetic changes throughout tumor development. For instance, tumor cells were found to harbor an overrepresentation of mutations in epigenetic writers, readers and erasers as well as chromatin remodeling complexes [284]. For example, DNMT3A is frequently mutated in hematological malignancies driving enhancer hypomethylation [285]. In contrast,

mutations in IDH1, IDH2 and TET enzymes result in an hypermethylation profile [202, 286]. On the other hand, epigenetic alterations may also influence genetics. For example, DNA repair genes, such as MGMT are frequently hypermethylated and silenced in cancer leading to increased mutation rates. It is already known for decades that all human cancers are associated with wide-spread epigenetic histone and/or DNA methylation changes [287, 288]. This is not surprising as a cancer cell has a complete different cellular identity compared to their normal counterpart. Although tumor-associated epigenetic patterns may be highly heterogenous, in general, tumor cells are characterized by a global DNA hypomethylation pattern, together with focal hypermethylation of mainly CGIs (**Figure 9**).

It was first suggested that hypermethylation of promoter CGIs leads to the silencing of tumor suppressor genes and consequently contribute to cancer. Indeed, some known tumor suppressor genes were found to be silenced by DNA hypermethylation in cancer, including RB, VHL, p16, MLH1 and BRCA1. However, later it became clear that most of the methylated genes are already repressed in normal cells [289]. In addition, hypermethylated genes are strongly enriched for the polycomb repressive H3K27me3 histone mark [290]. The EZH2 protein, which is the enzymatic component of the polycomb complex, is able to recruit DNMTs resulting in *de novo* methylation in cancer cells [225]. These results therefore imply that *de novo* methylation in cancer cells is the result of a highly regulated and instructive process. Polycomb repression is important in stem cells to reversible silence genes needed for differentiation. During differentiation these genes are active and become again silenced in adult cells. It is therefore hypothesized that DNA methylation leads to the permanent silencing of these genes, locking the cell in a stem cell-like self-renewal state which may predispose them for cancer formation [291].

Initial studies were mainly focused on CGIs and gene promoters using targeted approaches, however, with the emergency of genome-wide DNA methylation platforms we now have a much broader and better picture of healthy and cancer epigenomes. From these studies it became clear that hypermethylated changes are not exclusively found in CGIs, and that often more DNA methylation alterations are present in CpG-poor distally regulatory regions [37]. For example, using whole genome bisulfite sequencing of samples from different cancers, mainly super-enhancers were found to show strong DNA methylation shifts [292, 293]. In multiple myeloma, hypermethylation was primary localized in B-cell-specific TFBSs and not in CGI promoter regions [294]. DNA methylation alterations in colon cancer were mainly found in regions just next to CGIs, which were called CGI shores [37]. These studies also demonstrated that epigenetic profiles are highly heterogeneous, and that global loss and focal gains of DNA methylation is a hallmark of cancer epigenomes.

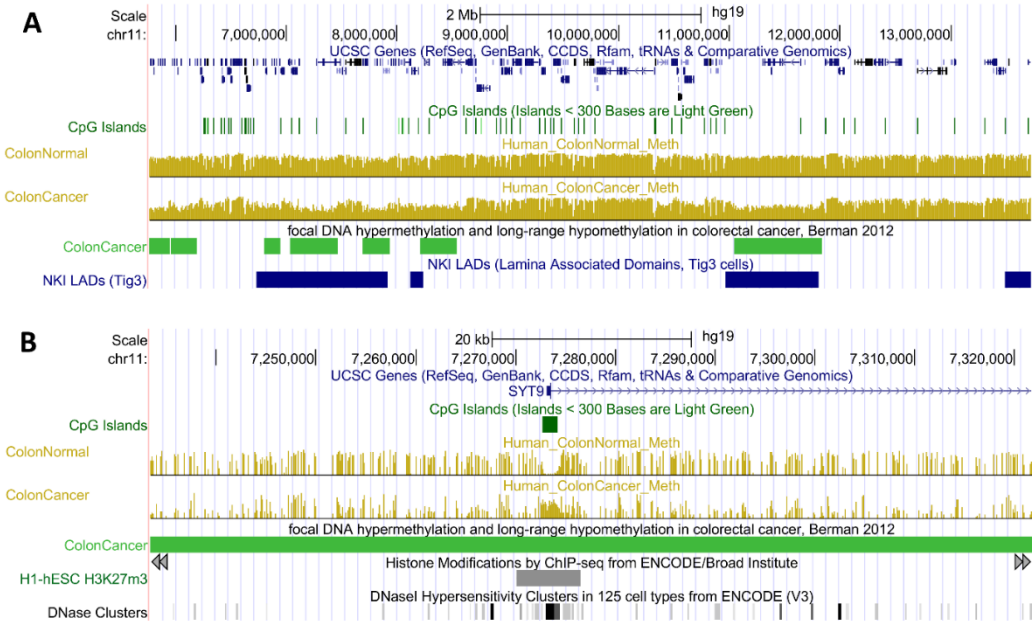


Figure 9: DNA methylation profile in cancer. Comparison of DNA methylation from colon cancer samples with normal colon tissues reveals **A)** large hypomethylated blocks (green bars) largely overlapping with LADs (dark blue bars) and **B)** focal hypermethylated regions overlapping with CGIs and H3K27me3 histone marks. This figure was created using the UCSC genome browser and bisulfite sequencing data from [297]. The height of the yellow bars represents the average DNA methylation values in the colon samples.

In many solid cancers, loss of DNA methylation in cancer occurs in large genomic blocks (5 kb – 10 Mb), which coincide with heterochromatic LOCKs, TADs and LADs, indicating the involvement of histone modifications and higher-order chromosomal organization [295-299]. It is also in these hypomethylated blocks where the focal hypermethylated regions reside leading to an erosion of the normal DNA methylation profile [297, 298]. Interestingly, loss of DNA methylation seems to be an early event in cancer formation and progress during the tumorigenesis, with metastasis having lower methylation levels than primary tumors [292]. Because of its known repressive effect on transcription, a global loss of DNA methylation lead to regional conversion of heterochromatin into euchromatin resulting in stochastic gene expression in these regions. As described previously, methylation of repeat elements are important in transposon repression and maintaining genomic stability. Consequently, it has been shown that global DNA hypomethylation promotes chromosomal instability and tumor formation in mice models and human cancer cell lines [269-272]. In another study, abnormal hypomethylation coincides with somatic copy number alterations in cancers [300]. Hypomethylation in tumors also leads to the activation of transposons resulting in increased mutational rate [301, 302]. Another consequence of global

hypomethylation is the reactivation of genes which were silenced by DNA methylation [303]. Many of these genes are germ-line specific genes as they are only expressed during germ line formation, and may contribute to tumor development by promoting oncogenic pathways.

Based on the improved genome-wide DNA methylation mapping technologies, we now have a clear picture of the extensive DNA methylation changes occurring in cancer cells. However, what is less clear is which methylation changes contribute to tumorigenesis and which changes are just bystanders and rather a consequence of the malignancy process [304, 305]. Just as in cancer genetics, some epimutations are “drivers” which will induce oncogenic pathways, while others are “passengers” and are not able to initiate or progress tumorigenesis. Identifying this driver and passenger epimutations may be important for therapeutic approaches. Potential driver epimutations are the ones which are located in known tumor suppressor genes, such as BRCA1, CDKN2A and MLH1. These genes have crucial roles in control of DNA repair and cell division and it may be expected that epigenetically silencing of these genes may contribute to tumor development. We have seen that most of the hypermethylated CGIs in cancer are polycomb repressed in embryonic stem cells leading to a restrictive cellular state with stem-cell renewal properties. On the other hand, epigenetic plasticity may also contribute to tumorigenesis [304, 306-308]. Some stimuli, including genetic and environmental factors, may induce permissive chromatin states leading to epigenetic instability or plasticity and subsequently large heterogeneity of chromatin and transcriptional states. Some of these states may harbor an increased fitness over the other chromatin states leading to the selection and clonal expansion of specific cancer cells. A clear example is the gain-of-function mutation in the IDH gene seen in glioma and leukemia leading to the inhibition of TET enzymes and overall hypermethylation. Stochastic hypermethylation events may disrupt the binding of CTCF which act as an insulator regulating gene-enhancer interactions. Disruption of CTCF binding may lead to the aberrant activation of nearby genes. Only some of these methylation changes may confer a growth advantage leading to the selection of this clone, while others are just passenger epimutations. In gliomas this is often observed for PDGFRA, a known oncogene, which is aberrantly expressed after the disruption of an CTCF boundary.

In addition to genetic factors (e.g. IDH mutations), also non-genetic factors may influence chromatin and epigenetic plasticity leading to oncogenesis. For example aging, which is the main risk factor for many cancers, is also accompanied by methylation changes in polycomb repressed genes in stem cells [309] and global loss of DNA methylation [310]. The global hypomethylated blocks found in cancer were also detected with aging, and, more interesting, correlates with mitotic history [296, 311]. Tumor cells therefore exhibit extensive DNA hypomethylation because of their highly

proliferative nature. Other factors that may initiate tumor-associated DNA methylation changes are cellular metabolism, inflammation, oxidative stress, diet and gut microbiome. Epigenetic and DNA methylation changes may therefore by a very early event in tumor formation [312]. For example, epigenetic variability in normal cells could predict the risk of cervical neoplasia [313]. Early dysregulation of the epigenome, by genetic factors (mutations in IDH and other chromatin-regulating enzymes) or environmental factors (aging, inflammation, etc), may lead to epigenetic instability and plasticity resulting in higher epigenetic variability in cancer. Some of these stochastic epigenetic changes will be passengers while others will be drivers which will subsequently be selected. Consequently, same cancer types may manifest in different ways across individuals and exhibit high epigenetic intratumoral heterogeneity which may have clinical consequences [314]. In addition, considerable DNA methylation changes can occur across primary, recurrent and metastatic tumors [315].

DNA methylation in complex non-malignant diseases

Risk of complex non-malignant diseases, including CVDs, diabetes, and neurological diseases, is modulated by a complex interplay of genetic, environmental and lifestyle factors. In the beginning of the 2000s, GWAS identified a lot of genetic risk factors [2], but it became increasingly clear that these SNPs could only explain a small part of disease heritability leaving a considerable part of the disease risk variability unexplained [4]. Epigenetic mechanisms may fill partially this gap of missing heritability. As we have seen previously, environmental and lifestyle factors may influence epigenetic patterns, and it is especially the prenatal and early developmental life periods which are highly sensitive towards environmental exposures leading to complex diseases later in adult life (DOHaD hypothesis) [5-7]. It has therefore been proposed that epigenetics could act as mediator between environment and disease [11]. Epigenetic mechanisms translate environmental exposures into adaptive transcriptional profiles which may contribute to health and disease. To better understand the epigenetic basis of complex diseases and traits, recently, many EWAS have been performed in healthy and diseased human populations [316] and found methylation at many CpG sites to be associated with almost all complex diseases, including CVDs [317], obesity [318], diabetes [319], Alzheimer's disease [320], Parkinson's disease [321], multiple sclerosis [322], asthma [323], etc. Therefore, just as in cancer epigenetics may play important roles in disease etiology and/or progression of common complex diseases.

In tumor cells, genes are activated or repressed by either complete demethylation or methylation of genomic regulatory regions. In contrast, in many complex non-

malignant disorders and phenotypes the DNA methylation changes are much more subtle, in the range of 1-10%, and often comprise single CpG sites [324]. The biological function of these small changes is still questionable, and whether it has direct transcriptional effects is often not clear. Another issue is the causality of the changes. Is the methylation change already present before disease onset, and initiating the disease process, or is the methylation change a consequence of the disease development? Because of the observational nature of most of the EWAS, it is difficult to answer this question, and longitudinal studies will be needed to give a definitive answer. Some methods exists which use genetic information to infer causality in EWAS [325]. In this way it was, for example, shown that most DNA methylation changes are a consequence of obesity and not a cause [326]. Although DNA methylation changes have been associated with numbers of environmental exposures on the one hand, and with diseases on the other hand, it is not clear yet whether DNA methylation is really a mediator between the environment, genetics and disease. Another issue is the tissue of interest, which is not always available. A lot of EWAS used whole blood samples, which may not always be the most relevant tissue of interest to resolve disease associated adverse epigenetic mechanisms. For now, it is not always clear whether DNA methylation patterns identified in blood samples can be applied as surrogate disease biomarkers for other tissues. Furthermore, blood samples and other tissues are complex mixtures of cells compromising different cell types, each with their own DNA methylation profile. Therefore DNA methylation changes could also be the result of shifts in cell type distribution, and not because of real intrinsic DNA methylation changes. Statistical tools exist, making use of reference methylomes of purified cell types, to estimate cell type composition which may be used to correct for these cellular effects [327]. However, not for all cell types reference methylomes are yet available.

DNA methylation based clinical applications

DNA methylation as a biomarker

Multiple commercially available genetic tests exist to assist in disease diagnosis, prognosis and therapy response, and may even inform healthy individuals about their risk profile for different common diseases, including neurological and metabolic disorders [3]. A disadvantage of these markers is that it cannot capture the contribution of environmental and lifestyle factors which often forms complex interactions with

genetic polymorphisms. Therefore, it is believed that epigenetic biomarkers may further help in disease risk prediction as they are able to provide information about past environmental exposures. In addition, DNA methylation marks are not static and aberrant DNA methylation marks can be reversed. Therefore, DNA methylation may also be interesting as a dynamic health marker evaluating the response success of therapeutic and/or lifestyle interventions. Other advantages of DNA-based biomarkers over RNA- or protein-based biomarkers, are their stability and the low amount of sample needed. Furthermore, multiple highly quantitatively and cost-efficient methods for measuring DNA methylation exists [328]. DNA methylation can be measured in non-invasive liquid biopsies, such as blood, saliva, urine and semen. Even circulating cell-free DNA (cfDNA) present in plasma can be used to detect abnormal DNA methylation profiles, especially for cancer applications [329, 330]. Interestingly, a recent study demonstrated the use of cfDNA methylation in identifying the tissue of origin which can be used in the diagnosis of multiple other diseases [331].

Today, only a few commercially available epi-biomarkers exist and they are all in the oncology field [332]. BMP3, NDRG4 and VIM hypermethylation measured in stool can be used as an early diagnostic marker for colorectal cancer (CRC) [333, 334]. Also a test in plasma where SEPT9 methylation could help in CRC diagnosis exists [335, 336]. However, mainly because of their lower specificity and higher cost as compared to traditional CRC screening methods, their use in clinical settings is limited. Other tests exist which could help (often in combination with other tests) in the diagnosis of lung, prostate and bladder cancer [332]. Since MGMT promoter methylation and silencing is associated with increased tumor response to alkylating agents in glioblastoma, MGMT promoter methylation based pharmaco-epigenomic commercial tests have been developed to predict drug response using either primary tumor samples or peripheral blood [337, 338]. Preclinical studies also demonstrated proof of concept for the use of DNA methylation gene panels in identifying cancer subtypes, however, for now this didn't result in clinical commercial tests [332]. Identifying markers for cancer subtypes and therapy response may help in personalized treatments and precision medicine. DNA methylation sequence panels based on the Illumina genome-wide DNA methylation arrays have also resulted in a test to identify the tumor of origin in patients with metastatic carcinoma of unknown primary (CUP) cancer [339]. Here, the methylation profiles of the unknown metastatic samples are compared to the methylation profiles of known cancer reference methylomes, which may help identifying the correct primary origin of these tumors and result in better treatment choices.

As described in the previous section, also complex diseases are associated with DNA methylation abnormalities and multiple studies indeed identified several potential

biomarkers, although they have not yet been translated into a validated clinical test [332]. In several diseases, including neurodegenerative diseases, the target tissue is not accessible and therefore one should use easy-accessible surrogate tissues, including blood, saliva, serum or cerebrospinal fluid, to evaluate disease states. An example is Alzheimer's disease, where post-mortem brain tissues indeed showed extensive DNA methylation remodeling, but the concordance with other more non-invasive tissues is not yet clear [340]. Metabolic disorders are also linked with DNA methylation changes in adipose tissue, pancreas, skeletal muscle and blood [341]. Lifestyle factors have a strong impact on metabolic disease risk and may also influence DNA methylation profiles. For example, physical exercise leads to DNA methylation changes in adipose tissues [342]. It would therefore be interesting to study to what extent lifestyle factors can adjust adverse DNA methylation markers and whether DNA methylation changes can be used as a read-out to monitor efficacy of lifestyle interventions.

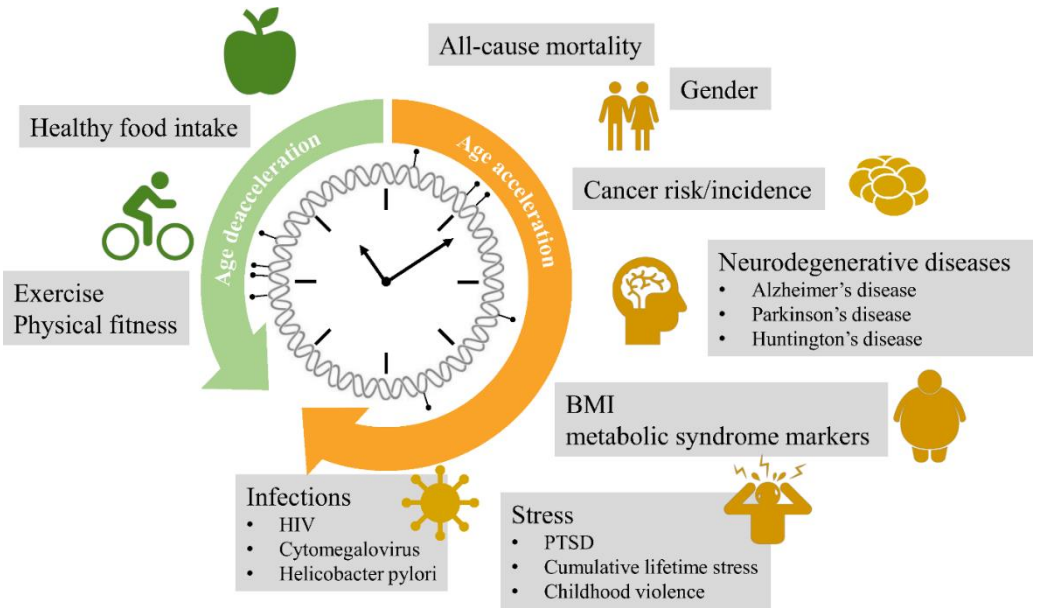


Figure 10: The epigenetic clock as a biological age predictor. Disease risk and lifestyle factors are shown known to be associated with either epigenetic age acceleration (orange) or deceleration (green). Figure was adapted from [138].

The possible use of DNA methylation as a health biomarker, was recently demonstrated by Steve Horvath and his epigenetic clock [139]. Using a large collection of DNA methylation datasets, he designed a multi-tissue estimator for chronological age. Subsequently, multiple studies demonstrated the use of this age predictor as a measure of biological age, whereas age acceleration (epigenetic age is larger than chronological age) is correlated with all-cause mortality, various age-related disorders and

phenotypes and age deceleration (epigenetic age is lower than chronological age) demonstrated lower risk for age-related disorders (**Figure 10**) [138, 343]. Interestingly, various healthy lifestyle factors, including fish intake, moderate alcohol consumption, education and fruit/vegetables intake are associated with epigenetic age deceleration [344]. Also caloric restriction which is known to extend lifespan in mammals is able to delay epigenetic aging [345]. In contrast, other factors, including stress [346] and physical inactivity [347] could increase epigenetic age acceleration. Of note, this epigenetic clock is now also commercially available provided by Zymo Research.

To summarize, although DNA methylation is promising as biomarker for various complex diseases, still some challenges must be overcome to translate preclinical studies into validated commercial clinical tests. The recently developed "epigenetic clock" algorithm has proven that DNA methylation can measure health status and can be used to monitor lifestyle interventions, extending clinical epigenetics beyond cancer applications.

DNA methylation as a target for therapeutic interventions

Given the extensive epigenetic abnormalities in cancer, epigenetic marks form an interesting target for therapeutic interventions. The first epi-drugs on the market were DNMT inhibitors (DNMTi), consisting of the nucleoside analogs azacytidine (5-AZA) and its deoxy derivative decitabine (5-Aza-CdR) [348]. They are FDA-approved and in clinical use for the treatment of myelodysplastic syndrome (MDS), acute myeloid leukemia (AML) and chronic myelomonocytic leukemia (CML). Although for certain patients these drugs can be highly effective, other patients don't respond and resistance to these therapies is common. In addition, their short half-life times hamper their use in solid tumors [349]. For DNMTi to be effective, cells should be in S phase at the time of exposure. Other epi-drugs include HDAC inhibitors (HDACi) which are also used for hematological cancers, including refractory cutaneous T cell lymphoma (CTCL), peripheral T cell lymphoma (PTCL) and multiple myeloma, often in combination with other drugs [350]. A different class of epi-drugs are iBETs which bind to the bromodomains of the BET proteins. BET proteins are readers of acetylated histones, and inhibition of these BET proteins may block expression of certain oncogenes, such as MYC [351]. iBETs are currently being tested in clinical trials.

All of the above mentioned drugs have genome-wide epigenetic effects, due to their general aspecific mechanism of inhibition. More specific targeted epi-therapies are now also being investigated in (pre)clinical trials. For example, inhibition of the histone methyltransferase EZH2 has cell death effects only in cells with EZH2 activating mutations [352]. Similarly, IDH inhibitors may selectively affect IDH mutant

malignancies [353]. Inhibition of the H3K79 methyltransferase DOT1L is only effective in leukemias with activation of MLL [354]. Another example, is the inhibition of the lysine histone demethylase LSD1, which is only effective in small-cell lung carcinoma (SCLC) with a specific DNA hypomethylation profile [355]. Currently, a lot of these epi-drugs are now being tested in clinical trials, with the hope that new epi-drugs will be approved in the near future [18, 332].

DNMTi has hypomethylated action at low doses, but can also be directly cytotoxic at high doses by their direct incorporation into DNA and/or RNA [349]. High doses, however, lead to extensive side effects, making the low-dose hypomethylated action of DNMTi more promising. Hypomethylation of driver epimutations, such as hypermethylated silenced tumor suppressor genes could result in the reactivation of these genes and reduce tumor growth [18, 332]. On the other hand, inhibition of DNA methylation may also lead to the expression of genes which are normally silenced, such as cancer testis antigens and repetitive elements [356]. This activation may lead to the formation of neoantigens in these cells and thus increasing the immunogenicity by making the cancer cell more visible to the host immune defense mechanism [357]. For example, activation of endogenous retroviruses (ERVs) can lead to a state of viral mimicry, where dsRNAs are recognized by pattern recognition receptors leading to an activated interferon signaling response. Interferon may increase antigen presentation in the cancer cells resulting in increased visibility of the cancer cell to the adaptive immune system [357]. Therefore, combination therapy of DNMTi and immunotherapy is a promising new approach which may enhance the utility of DNMTi in solid tissues.

Epigenetic changes in other complex non-malignant diseases are much more subtle, and therefore epi-therapy has mainly focused on oncological applications. The use of general DNMT/HDAC inhibitors in the treatment of common non-malignant diseases is questionable given the rather modest adverse epigenetic changes. Furthermore, causality of most of these disease-associated DNA methylation changes is not necessary proven: we don't know whether a change is a consequence of the disease process or whether it may cause the disease. Non-causal DNA methylation changes may still be valuable as biomarker, but are probably not very useful as therapeutic target. Identifying the driver epi-mutations may therefore by crucial in selectively targeting specific genomic regions for epigenetic editing. For this, the CRISPR-Cas9 genome editing tool has now also been exploited for epigenetic purposes, by fusing the nuclease-deactivated Cas9 protein to an epigenome-modifying enzyme [358]. A guide RNA directs the enzyme to a genomic region of interest to change a particular epigenetic mark. This tool can therefore be used to selectively modify specific DNA methylation sites and may not only give information about the functionality of the site but will also be very promising as new form of epi-therapy.

References

1. Lander, E.S., et al., *Initial sequencing and analysis of the human genome.* Nature, 2001. **409**(6822): p. 860-921.
2. Hirschhorn, J.N. and M.J. Daly, *Genome-wide association studies for common diseases and complex traits.* Nat Rev Genet, 2005. **6**(2): p. 95-108.
3. Carrasco-Ramiro, F., R. Peiro-Pastor, and B. Aguado, *Human genomics projects and precision medicine.* Gene Ther, 2017. **24**(9): p. 551-561.
4. Manolio, T.A., et al., *Finding the missing heritability of complex diseases.* Nature, 2009. **461**(7265): p. 747-53.
5. Barker, D.J., *The origins of the developmental origins theory.* J Intern Med, 2007. **261**(5): p. 412-7.
6. Yamada, L. and S. Chong, *Epigenetic studies in Developmental Origins of Health and Disease: pitfalls and key considerations for study design and interpretation.* J Dev Orig Health Dis, 2017. **8**(1): p. 30-43.
7. Bianco-Miotto, T., et al., *Epigenetics and DOHaD: from basics to birth and beyond.* J Dev Orig Health Dis, 2017. **8**(5): p. 513-519.
8. Allis, C.D. and T. Jenuwein, *The molecular hallmarks of epigenetic control.* Nat Rev Genet, 2016. **17**(8): p. 487-500.
9. Feil, R. and M.F. Fraga, *Epigenetics and the environment: emerging patterns and implications.* Nat Rev Genet, 2012. **13**(2): p. 97-109.
10. Henikoff, S. and J.M. Greally, *Epigenetics, cellular memory and gene regulation.* Curr Biol, 2016. **26**(14): p. R644-8.
11. Kim, M. and J. Costello, *DNA methylation: an epigenetic mark of cellular memory.* Exp Mol Med, 2017. **49**(4): p. e322.
12. Haberle, V. and A. Stark, *Eukaryotic core promoters and the functional basis of transcription initiation.* Nat Rev Mol Cell Biol, 2018. **19**(10): p. 621-637.
13. Chen, F.X., E.R. Smith, and A. Shilatifard, *Born to run: control of transcription elongation by RNA polymerase II.* Nat Rev Mol Cell Biol, 2018. **19**(7): p. 464-478.
14. Li, B., M. Carey, and J.L. Workman, *The role of chromatin during transcription.* Cell, 2007. **128**(4): p. 707-19.
15. Soshnev, A.A., S.Z. Josefowicz, and C.D. Allis, *Greater Than the Sum of Parts: Complexity of the Dynamic Epigenome.* Mol Cell, 2016. **62**(5): p. 681-94.
16. Kinde, B., et al., *Reading the unique DNA methylation landscape of the brain: Non-CpG methylation, hydroxymethylation, and MeCP2.* Proc Natl Acad Sci U S A, 2015. **112**(22): p. 6800-6.
17. Pinney, S.E., *Mammalian Non-CpG Methylation: Stem Cells and Beyond.* Biology (Basel), 2014. **3**(4): p. 739-51.
18. Jones, P.A., J.P. Issa, and S. Baylin, *Targeting the cancer epigenome for therapy.* Nat Rev Genet, 2016. **17**(10): p. 630-41.
19. Esteller, M., *Non-coding RNAs in human disease.* Nat Rev Genet, 2011. **12**(12): p. 861-74.
20. Pombo, A. and N. Dillon, *Three-dimensional genome architecture: players and mechanisms.* Nat Rev Mol Cell Biol, 2015. **16**(4): p. 245-57.
21. Rowley, M.J. and V.G. Corces, *Organizational principles of 3D genome architecture.* Nat Rev Genet, 2018. **19**(12): p. 789-800.
22. Bonev, B. and G. Cavalli, *Organization and function of the 3D genome.* Nat Rev Genet, 2016. **17**(12): p. 772.
23. Heinz, S., et al., *The selection and function of cell type-specific enhancers.* Nat Rev Mol Cell Biol, 2015. **16**(3): p. 144-54.
24. Zaret, K.S. and S.E. Mango, *Pioneer transcription factors, chromatin dynamics, and cell fate control.* Curr Opin Genet Dev, 2016. **37**: p. 76-81.
25. Takahashi, K. and S. Yamanaka, *Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors.* Cell, 2006. **126**(4): p. 663-76.

26. Srivastava, D. and N. DeWitt, *In Vivo Cellular Reprogramming: The Next Generation.* Cell, 2016. **166**(6): p. 1386-1396.
27. Zhou, Y., et al., *Bmi1 Is a Key Epigenetic Barrier to Direct Cardiac Reprogramming.* Cell Stem Cell, 2016. **18**(3): p. 382-95.
28. Miles, D.C., et al., *TRIM28 is an Epigenetic Barrier to Induced Pluripotent Stem Cell Reprogramming.* Stem Cells, 2017. **35**(1): p. 147-157.
29. Chen, J., et al., *H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs.* Nat Genet, 2013. **45**(1): p. 34-42.
30. Lister, R., et al., *Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells.* Nature, 2011. **471**(7336): p. 68-73.
31. Nashun, B., P.W. Hill, and P. Hajkova, *Reprogramming of cell fate: epigenetic memory and the erasure of memories past.* EMBO J, 2015. **34**(10): p. 1296-308.
32. Ehrlich, M., X.Y. Zhang, and N.M. Inamdar, *Spontaneous deamination of cytosine and 5-methylcytosine residues in DNA and replacement of 5-methylcytosine residues with cytosine residues.* Mutat Res, 1990. **238**(3): p. 277-86.
33. Antonarakis, S.E., M. Krawczak, and D.N. Cooper, *Disease-causing mutations in the human genome.* Eur J Pediatr, 2000. **159 Suppl 3**: p. S173-8.
34. Cooper, D.N. and H. Youssoufian, *The CpG dinucleotide and human genetic disease.* Hum Genet, 1988. **78**(2): p. 151-5.
35. Illingworth, R.S. and A.P. Bird, *CpG islands--'a rough guide'.* FEBS Lett, 2009. **583**(11): p. 1713-20.
36. Deaton, A.M. and A. Bird, *CpG islands and the regulation of transcription.* Genes Dev, 2011. **25**(10): p. 1010-22.
37. Irizarry, R.A., et al., *The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores.* Nat Genet, 2009. **41**(2): p. 178-186.
38. Cantone, I. and A.G. Fisher, *Epigenetic programming and reprogramming during development.* Nat Struct Mol Biol, 2013. **20**(3): p. 282-9.
39. Smith, Z.D. and A. Meissner, *DNA methylation: roles in mammalian development.* Nat Rev Genet, 2013. **14**(3): p. 204-20.
40. Santos, F., et al., *Dynamic reprogramming of DNA methylation in the early mouse embryo.* Dev Biol, 2002. **241**(1): p. 172-82.
41. Oswald, J., et al., *Active demethylation of the paternal genome in the mouse zygote.* Curr Biol, 2000. **10**(8): p. 475-8.
42. Wu, X. and Y. Zhang, *TET-mediated active DNA demethylation: mechanism, function and beyond.* Nat Rev Genet, 2017. **18**(9): p. 517-534.
43. Iqbal, K., et al., *Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine.* Proc Natl Acad Sci U S A, 2011. **108**(9): p. 3642-7.
44. Gu, T.P., et al., *The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes.* Nature, 2011. **477**(7366): p. 606-10.
45. Inoue, A., et al., *Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development.* Cell Res, 2011. **21**(12): p. 1670-6.
46. Inoue, A. and Y. Zhang, *Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos.* Science, 2011. **334**(6053): p. 194.
47. Shen, L., et al., *Tet3 and DNA replication mediate demethylation of both the maternal and paternal genomes in mouse zygotes.* Cell Stem Cell, 2014. **15**(4): p. 459-471.
48. Guo, F., et al., *Active and passive demethylation of male and female pronuclear DNA in the mammalian zygote.* Cell Stem Cell, 2014. **15**(4): p. 447-459.
49. Amouroux, R., et al., *De novo DNA methylation drives 5hmC accumulation in mouse zygotes.* Nat Cell Biol, 2016. **18**(2): p. 225-233.
50. von Meyenn, F., et al., *Impairment of DNA Methylation Maintenance Is the Main Cause of Global Demethylation in Naive Embryonic Stem Cells.* Mol Cell, 2016. **62**(6): p. 848-861.
51. Okano, M., et al., *DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development.* Cell, 1999. **99**(3): p. 247-57.

52. Ooi, S.K., et al., *DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA.* Nature, 2007. **448**(7154): p. 714-7.
53. Otani, J., et al., *Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain.* EMBO Rep, 2009. **10**(11): p. 1235-41.
54. Gowher, H., et al., *Mechanism of stimulation of catalytic activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5)-methyltransferases by Dnmt3L.* J Biol Chem, 2005. **280**(14): p. 13341-8.
55. Suetake, I., et al., *DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction.* J Biol Chem, 2004. **279**(26): p. 27816-23.
56. Guo, X., et al., *Structural insight into autoinhibition and histone H3-induced activation of DNMT3A.* Nature, 2015. **517**(7536): p. 640-4.
57. Brandeis, M., et al., *Sp1 elements protect a CpG island from de novo methylation.* Nature, 1994. **371**(6496): p. 435-8.
58. Greenfield, R., et al., *Role of transcription complexes in the formation of the basal methylation pattern in early development.* Proc Natl Acad Sci U S A, 2018. **115**(41): p. 10387-10391.
59. Long, H.K., N.P. Blackledge, and R.J. Klose, *ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection.* Biochem Soc Trans, 2013. **41**(3): p. 727-40.
60. Blackledge, N.P., J.P. Thomson, and P.J. Skene, *CpG island chromatin is shaped by recruitment of ZF-CxxC proteins.* Cold Spring Harb Perspect Biol, 2013. **5**(11): p. a018648.
61. Clouaire, T., et al., *Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells.* Genes Dev, 2012. **26**(15): p. 1714-28.
62. Thomson, J.P., et al., *CpG islands influence chromatin structure via the CpG-binding protein Cfp1.* Nature, 2010. **464**(7291): p. 1082-6.
63. Song, J., et al., *Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation.* Science, 2012. **335**(6069): p. 709-12.
64. Sharif, J., et al., *The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA.* Nature, 2007. **450**(7171): p. 908-12.
65. Bostick, M., et al., *UHRF1 plays a role in maintaining DNA methylation in mammalian cells.* Science, 2007. **317**(5845): p. 1760-4.
66. Fatemi, M., et al., *Dnmt3a and Dnmt1 functionally cooperate during de novo methylation of DNA.* Eur J Biochem, 2002. **269**(20): p. 4981-4.
67. Li, Y., et al., *Stella safeguards the oocyte methylome by preventing de novo methylation mediated by DNMT1.* Nature, 2018. **564**(7734): p. 136-140.
68. Jeltsch, A. and R.Z. Jurkowska, *New concepts in DNA methylation.* Trends Biochem Sci, 2014. **39**(7): p. 310-8.
69. Liang, G., et al., *Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements.* Mol Cell Biol, 2002. **22**(2): p. 480-91.
70. Williams, K., et al., *TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity.* Nature, 2011. **473**(7347): p. 343-8.
71. Nishiyama, A., et al., *Uhrf1-dependent H3K23 ubiquitylation couples maintenance DNA methylation and replication.* Nature, 2013. **502**(7470): p. 249-53.
72. Zhao, Q., et al., *Dissecting the precise role of H3K9 methylation in crosstalk with DNA maintenance methylation in mammals.* Nat Commun, 2016. **7**: p. 12464.
73. Ziller, M.J., et al., *Charting a dynamic DNA methylation landscape of the human genome.* Nature, 2013. **500**(7463): p. 477-81.
74. Feldman, N., et al., *G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis.* Nat Cell Biol, 2006. **8**(2): p. 188-94.
75. Bergman, Y. and H. Cedar, *DNA methylation dynamics in health and disease.* Nat Struct Mol Biol, 2013. **20**(3): p. 274-81.
76. Epsztejn-Litman, S., et al., *De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes.* Nat Struct Mol Biol, 2008. **15**(11): p. 1176-1183.
77. Dawlaty, M.M., et al., *Loss of Tet enzymes compromises proper differentiation of embryonic stem cells.* Dev Cell, 2014. **29**(1): p. 102-11.
78. Li, C., et al., *Overlapping Requirements for Tet2 and Tet3 in Normal Development and Hematopoietic Stem Cell Emergence.* Cell Rep, 2015. **12**(7): p. 1133-43.

79. Seritrakul, P. and J.M. Gross, *Tet-mediated DNA hydroxymethylation regulates retinal neurogenesis by modulating cell-extrinsic signaling pathways.* PLoS Genet, 2017. **13**(9): p. e1006987.
80. Lio, C.W., et al., *Tet2 and Tet3 cooperate with B-lineage transcription factors to regulate DNA modification and chromatin accessibility.* Elife, 2016. **5**.
81. Orlanski, S., et al., *Tissue-specific DNA demethylation is required for proper B-cell differentiation and function.* Proc Natl Acad Sci U S A, 2016. **113**(18): p. 5018-23.
82. de la Rica, L., et al., *PU.1 target genes undergo Tet2-coupled demethylation and DNMT3b-mediated methylation in monocyte-to-osteoclast differentiation.* Genome Biol, 2013. **14**(9): p. R99.
83. Suzuki, T., et al., *RUNX1 regulates site specificity of DNA demethylation by recruitment of DNA demethylation machineries in hematopoietic cells.* Blood Adv, 2017. **1**(20): p. 1699-1711.
84. Marchal, C. and B. Miotto, *Emerging concept in DNA methylation: role of transcription factors in shaping DNA methylation patterns.* J Cell Physiol, 2015. **230**(4): p. 743-51.
85. Slieker, R.C., et al., *Identification and systematic annotation of tissue-specific differentially methylated regions using the Illumina 450k array.* Epigenetics Chromatin, 2013. **6**(1): p. 26.
86. Gu, J., et al., *Mapping of Variable DNA Methylation Across Multiple Cell Types Defines a Dynamic Regulatory Landscape of the Human Genome.* G3 (Bethesda), 2016. **6**(4): p. 973-86.
87. Maunakea, A.K., et al., *Conserved role of intragenic DNA methylation in regulating alternative promoters.* Nature, 2010. **466**(7303): p. 253-7.
88. Deaton, A.M., et al., *Cell type-specific DNA methylation at intragenic CpG islands in the immune system.* Genome Res, 2011. **21**(7): p. 1074-86.
89. Guo, F., et al., *The Transcriptome and DNA Methylome Landscapes of Human Primordial Germ Cells.* Cell, 2015. **161**(6): p. 1437-52.
90. Tang, W.W., et al., *A Unique Gene Regulatory Network Resets the Human Germline Epigenome for Development.* Cell, 2015. **161**(6): p. 1453-67.
91. Kagiwada, S., et al., *Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice.* EMBO J, 2013. **32**(3): p. 340-53.
92. Hargan-Calvopina, J., et al., *Stage-Specific Demethylation in Primordial Germ Cells Safeguards against Precocious Differentiation.* Dev Cell, 2016. **39**(1): p. 75-86.
93. Seisenberger, S., et al., *The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells.* Mol Cell, 2012. **48**(6): p. 849-62.
94. Ohno, R., et al., *A replication-dependent passive mechanism modulates DNA demethylation in mouse primordial germ cells.* Development, 2013. **140**(14): p. 2892-903.
95. Yamaguchi, S., et al., *Dynamics of 5-methylcytosine and 5-hydroxymethylcytosine during germ cell reprogramming.* Cell Res, 2013. **23**(3): p. 329-39.
96. Hackett, J.A., et al., *Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine.* Science, 2013. **339**(6118): p. 448-52.
97. Kato, Y., et al., *Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse.* Hum Mol Genet, 2007. **16**(19): p. 2272-80.
98. Kaneda, M., et al., *Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting.* Nature, 2004. **429**(6994): p. 900-3.
99. Singh, P., et al., *De novo DNA methylation in the male germ line occurs by default but is excluded at sites of H3K4 methylation.* Cell Rep, 2013. **4**(1): p. 205-19.
100. Gahurova, L., et al., *Transcription and chromatin determinants of de novo DNA methylation timing in oocytes.* Epigenetics Chromatin, 2017. **10**: p. 25.
101. Nagamori, I., et al., *Comprehensive DNA Methylation Analysis of Retrotransposons in Male Germ Cells.* Cell Rep, 2015. **12**(10): p. 1541-7.
102. Monk, D., et al., *Genomic imprinting disorders: lessons on how genome, epigenome and environment interact.* Nat Rev Genet, 2019.
103. Busche, S., et al., *Population whole-genome bisulfite sequencing across two tissues highlights the environment as the principal source of human methylome variation.* Genome Biol, 2015. **16**: p. 290.

104. Garg, P., et al., *A survey of inter-individual variation in DNA methylation identifies environmentally responsive co-regulated networks of epigenetic variation in the human genome.* PLoS Genet, 2018. **14**(10): p. e1007707.
105. Chatterjee, A., et al., *Genome-wide DNA methylation map of human neutrophils reveals widespread inter-individual epigenetic variation.* Sci Rep, 2015. **5**: p. 17328.
106. Grundberg, E., et al., *Global analysis of DNA methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements.* Am J Hum Genet, 2013. **93**(5): p. 876-90.
107. Yet, I., et al., *Genetic and environmental impacts on DNA methylation levels in twins.* Epigenomics, 2016. **8**(1): p. 105-17.
108. Gordon, L., 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**(8): p. 1395-406.
109. Kaminsky, Z.A., et al., *DNA methylation profiles in monozygotic and dizygotic twins.* Nat Genet, 2009. **41**(2): p. 240-5.
110. McRae, A.F., et al., *Contribution of genetic variation to transgenerational inheritance of DNA methylation.* Genome Biol, 2014. **15**(5): p. R73.
111. van Dongen, J., et al., *Genetic and environmental influences interact with age and sex in shaping the human methylome.* Nat Commun, 2016. **7**: p. 11115.
112. Bell, J.T., et al., *Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population.* PLoS Genet, 2012. **8**(4): p. e1002629.
113. McRae, A.F., et al., *Identification of 55,000 Replicated DNA Methylation QTL.* Sci Rep, 2018. **8**(1): p. 17605.
114. Do, C., et al., *Genetic-epigenetic interactions in cis: a major focus in the post-GWAS era.* Genome Biol, 2017. **18**(1): p. 120.
115. Lemire, M., et al., *Long-range epigenetic regulation is conferred by genetic variation located at thousands of independent loci.* Nat Commun, 2015. **6**: p. 6326.
116. Pierce, B.L., et al., *Co-occurring expression and methylation QTLs allow detection of common causal variants and shared biological mechanisms.* Nat Commun, 2018. **9**(1): p. 804.
117. Shakhbazov, K., et al., *Shared genetic control of expression and methylation in peripheral blood.* BMC Genomics, 2016. **17**: p. 278.
118. Hannon, E., et al., *Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci.* Nat Neurosci, 2016. **19**(1): p. 48-54.
119. Gamazon, E.R., et al., *Enrichment of cis-regulatory gene expression SNPs and methylation quantitative trait loci among bipolar disorder susceptibility variants.* Mol Psychiatry, 2013. **18**(3): p. 340-6.
120. Volkov, P., et al., *A Genome-Wide mQTL Analysis in Human Adipose Tissue Identifies Genetic Variants Associated with DNA Methylation, Gene Expression and Metabolic Traits.* PLoS One, 2016. **11**(6): p. e0157776.
121. Gong, J., et al., *Pancan-meQTL: a database to systematically evaluate the effects of genetic variants on methylation in human cancer.* Nucleic Acids Res, 2018.
122. Wu, Y., et al., *Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits.* Nat Commun, 2018. **9**(1): p. 918.
123. Mancuso, N., et al., *Integrating Gene Expression with Summary Association Statistics to Identify Genes Associated with 30 Complex Traits.* Am J Hum Genet, 2017. **100**(3): p. 473-487.
124. Hannon, E., et al., *Pleiotropic Effects of Trait-Associated Genetic Variation on DNA Methylation: Utility for Refining GWAS Loci.* Am J Hum Genet, 2017. **100**(6): p. 954-959.
125. Hannon, E., et al., *Leveraging DNA-Methylation Quantitative-Trait Loci to Characterize the Relationship between Methylomic Variation, Gene Expression, and Complex Traits.* Am J Hum Genet, 2018. **103**(5): p. 654-665.
126. van Eijk, K.R., et al., *Genetic analysis of DNA methylation and gene expression levels in whole blood of healthy human subjects.* BMC Genomics, 2012. **13**: p. 636.
127. Gutierrez-Arcelus, M., et al., *Passive and active DNA methylation and the interplay with genetic variation in gene regulation.* Elife, 2013. **2**: p. e00523.

128. Banovich, N.E., et al., *Methylation QTLs are associated with coordinated changes in transcription factor binding, histone modifications, and gene expression levels.* PLoS Genet, 2014. **10**(9): p. e1004663.
129. Stadler, M.B., et al., *DNA-binding factors shape the mouse methylome at distal regulatory regions.* Nature, 2011. **480**(7378): p. 490-5.
130. Feldmann, A., et al., *Transcription factor occupancy can mediate active turnover of DNA methylation at regulatory regions.* PLoS Genet, 2013. **9**(12): p. e1003994.
131. Domcke, S., et al., *Competition between DNA methylation and transcription factors determines binding of NRF1.* Nature, 2015. **528**(7583): p. 575-9.
132. Pirazzini, C., et al., *Space/population and time/age in DNA methylation variability in humans: a study on IGF2/H19 locus in different Italian populations and in mono- and di-zygotic twins of different age.* Aging (Albany NY), 2012. **4**(7): p. 509-20.
133. Fraga, M.F., et al., *Epigenetic differences arise during the lifetime of monozygotic twins.* Proc Natl Acad Sci U S A, 2005. **102**(30): p. 10604-9.
134. Talens, R.P., et al., *Epigenetic variation during the adult lifespan: cross-sectional and longitudinal data on monozygotic twin pairs.* Aging Cell, 2012. **11**(4): p. 694-703.
135. Tan, Q., et al., *Epigenetic drift in the aging genome: a ten-year follow-up in an elderly twin cohort.* Int J Epidemiol, 2016. **45**(4): p. 1146-1158.
136. Wong, C.C., et al., *A longitudinal study of epigenetic variation in twins.* Epigenetics, 2010. **5**(6): p. 516-26.
137. Zheng, S.C., M. Widschwendter, and A.E. Teschendorff, *Epigenetic drift, epigenetic clocks and cancer risk.* Epigenomics, 2016. **8**(5): p. 705-19.
138. Declerck, K. and W. Vanden Berghe, *Back to the future: Epigenetic clock plasticity towards healthy aging.* Mech Ageing Dev, 2018. **174**: p. 18-29.
139. Horvath, S., *DNA methylation age of human tissues and cell types.* Genome Biol, 2013. **14**(10): p. R115.
140. Anderson, O.S., K.E. Sant, and D.C. Dolinoy, *Nutrition and epigenetics: an interplay of dietary methyl donors, one-carbon metabolism and DNA methylation.* J Nutr Biochem, 2012. **23**(8): p. 853-9.
141. Vinkers, C.H., et al., *Traumatic stress and human DNA methylation: a critical review.* Epigenomics, 2015. **7**(4): p. 593-608.
142. Alfano, R., et al., *The Impact of Air Pollution on Our Epigenome: How Far Is the Evidence? (A Systematic Review).* Curr Environ Health Rep, 2018. **5**(4): p. 544-578.
143. Harlaar, N. and K.E. Hutchison, *Alcohol and the methylome: design and analysis considerations for research using human samples.* Drug Alcohol Depend, 2013. **133**(2): p. 305-16.
144. Gao, X., et al., *DNA methylation changes of whole blood cells in response to active smoking exposure in adults: a systematic review of DNA methylation studies.* Clin Epigenetics, 2015. **7**: p. 113.
145. Zhang, Y., et al., *F2RL3 methylation, lung cancer incidence and mortality.* Int J Cancer, 2015. **137**(7): p. 1739-48.
146. Breitling, L.P., et al., *Smoking, F2RL3 methylation, and prognosis in stable coronary heart disease.* Eur Heart J, 2012. **33**(22): p. 2841-8.
147. Zhang, Y., et al., *F2RL3 methylation in blood DNA is a strong predictor of mortality.* Int J Epidemiol, 2014. **43**(4): p. 1215-25.
148. Zhang, Y., et al., *Smoking-Associated DNA Methylation Biomarkers and Their Predictive Value for All-Cause and Cardiovascular Mortality.* Environ Health Perspect, 2016. **124**(1): p. 67-74.
149. Zhang, Y., et al., *F2RL3 methylation as a biomarker of current and lifetime smoking exposures.* Environ Health Perspect, 2014. **122**(2): p. 131-7.
150. Zeilinger, S., et al., *Tobacco smoking leads to extensive genome-wide changes in DNA methylation.* PLoS One, 2013. **8**(5): p. e63812.
151. Guida, F., et al., *Dynamics of smoking-induced genome-wide methylation changes with time since smoking cessation.* Hum Mol Genet, 2015. **24**(8): p. 2349-59.
152. Tsaprouni, L.G., et al., *Cigarette smoking reduces DNA methylation levels at multiple genomic loci but the effect is partially reversible upon cessation.* Epigenetics, 2014. **9**(10): p. 1382-96.

153. Tsai, P.C., et al., *Smoking induces coordinated DNA methylation and gene expression changes in adipose tissue with consequences for metabolic health.* Clin Epigenetics, 2018. **10**(1): p. 126.
154. Widschwendter, M., et al., *Epigenome-based cancer risk prediction: rationale, opportunities and challenges.* Nat Rev Clin Oncol, 2018. **15**(5): p. 292-309.
155. Teh, A.L., et al., *The effect of genotype and in utero environment on interindividual variation in neonate DNA methylomes.* Genome Res, 2014. **24**(7): p. 1064-74.
156. Bui, M., et al., *Sharing a Placenta is Associated With a Greater Similarity in DNA Methylation in Monochorionic Versus Dichorionic Twin Pars in Blood at Age 14.* Twin Res Hum Genet, 2015. **18**(6): p. 680-5.
157. Ollikainen, M., et al., *DNA methylation analysis of multiple tissues from newborn twins reveals both genetic and intrauterine components to variation in the human neonatal epigenome.* Hum Mol Genet, 2010. **19**(21): p. 4176-88.
158. Felix, J.F., et al., *Cohort Profile: Pregnancy And Childhood Epigenetics (PACE) Consortium.* Int J Epidemiol, 2018. **47**(1): p. 22-23u.
159. Vaiserman, A., *Epidemiologic evidence for association between adverse environmental exposures in early life and epigenetic variation: a potential link to disease susceptibility?* Clin Epigenetics, 2015. **7**: p. 96.
160. Vickers, M.H., *Early life nutrition, epigenetics and programming of later life disease.* Nutrients, 2014. **6**(6): p. 2165-78.
161. Desai, M., J.K. Jellyman, and M.G. Ross, *Epigenomics, gestational programming and risk of metabolic syndrome.* Int J Obes (Lond), 2015. **39**(4): p. 633-41.
162. Barker, D.J., *The fetal and infant origins of adult disease.* BMJ, 1990. **301**(6761): p. 1111.
163. Barker, D.J. and C. Osmond, *Infant mortality, childhood nutrition, and ischaemic heart disease in England and Wales.* Lancet, 1986. **1**(8489): p. 1077-81.
164. O'Donnell, K.J. and M.J. Meaney, *Fetal Origins of Mental Health: The Developmental Origins of Health and Disease Hypothesis.* Am J Psychiatry, 2017. **174**(4): p. 319-328.
165. Oestreich, A.K. and K.H. Moley, *Developmental and Transmittable Origins of Obesity-Associated Health Disorders.* Trends Genet, 2017. **33**(6): p. 399-407.
166. Hales, C.N. and D.J. Barker, *The thrifty phenotype hypothesis.* Br Med Bull, 2001. **60**: p. 5-20.
167. Roseboom, T.J., et al., *Hungry in the womb: what are the consequences? Lessons from the Dutch famine.* Maturitas, 2011. **70**(2): p. 141-5.
168. Roseboom, T., S. de Rooij, and R. Painter, *The Dutch famine and its long-term consequences for adult health.* Early Hum Dev, 2006. **82**(8): p. 485-91.
169. Heijmans, B.T., et al., *Persistent epigenetic differences associated with prenatal exposure to famine in humans.* Proc Natl Acad Sci U S A, 2008. **105**(44): p. 17046-9.
170. Tobi, E.W., et al., *DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific.* Hum Mol Genet, 2009. **18**(21): p. 4046-53.
171. Tobi, E.W., et al., *DNA methylation signatures link prenatal famine exposure to growth and metabolism.* Nat Commun, 2014. **5**: p. 5592.
172. Tobi, E.W., et al., *DNA methylation as a mediator of the association between prenatal adversity and risk factors for metabolic disease in adulthood.* Sci Adv, 2018. **4**(1): p. eaao4364.
173. Tobi, E.W., et al., *Early gestation as the critical time-window for changes in the prenatal environment to affect the adult human blood methylome.* Int J Epidemiol, 2015. **44**(4): p. 1211-23.
174. Soubry, A., *Epigenetics as a Driver of Developmental Origins of Health and Disease: Did We Forget the Fathers?* Bioessays, 2018. **40**(1).
175. Bygren, L.O., G. Kaati, and S. Edvinsson, *Longevity determined by paternal ancestors' nutrition during their slow growth period.* Acta Biotheor, 2001. **49**(1): p. 53-9.
176. Kaati, G., L.O. Bygren, and S. Edvinsson, *Cardiovascular and diabetes mortality determined by nutrition during parents' and grandparents' slow growth period.* Eur J Hum Genet, 2002. **10**(11): p. 682-8.
177. Pembrey, M.E., et al., *Sex-specific, male-line transgenerational responses in humans.* Eur J Hum Genet, 2006. **14**(2): p. 159-66.

178. Vagero, D., et al., *Paternal grandfather's access to food predicts all-cause and cancer mortality in grandsons.* Nat Commun, 2018. **9**(1): p. 5124.
179. Northstone, K., et al., *Prepubertal start of father's smoking and increased body fat in his sons: further characterisation of paternal transgenerational responses.* Eur J Hum Genet, 2014. **22**(12): p. 1382-6.
180. Soubry, A., et al., *Obesity-related DNA methylation at imprinted genes in human sperm: Results from the TIEGER study.* Clin Epigenetics, 2016. **8**: p. 51.
181. Soubry, A., et al., *Newborns of obese parents have altered DNA methylation patterns at imprinted genes.* Int J Obes (Lond), 2015. **39**(4): p. 650-7.
182. Okae, H., et al., *Genome-wide analysis of DNA methylation dynamics during early human development.* PLoS Genet, 2014. **10**(12): p. e1004868.
183. Yehuda, R., et al., *Holocaust Exposure Induced Intergenerational Effects on FKBP5 Methylation.* Biol Psychiatry, 2016. **80**(5): p. 372-80.
184. Skvortsova, K., N. Iovino, and O. Bogdanovic, *Functions and mechanisms of epigenetic inheritance in animals.* Nat Rev Mol Cell Biol, 2018. **19**(12): p. 774-790.
185. Anway, M.D., et al., *Epigenetic transgenerational actions of endocrine disruptors and male fertility.* Science, 2005. **308**(5727): p. 1466-9.
186. Beck, D., I. Sadler-Riggleman, and M.K. Skinner, *Generational comparisons (F1 versus F3) of vinclozolin induced epigenetic transgenerational inheritance of sperm differential DNA methylation regions (epimutations) using MeDIP-Seq.* Environ Epigenet, 2017. **3**(3).
187. Iqbal, K., et al., *Deleterious effects of endocrine disruptors are corrected in the mammalian germline by epigenome reprogramming.* Genome Biol, 2015. **16**: p. 59.
188. Radford, E.J., *Exploring the extent and scope of epigenetic inheritance.* Nat Rev Endocrinol, 2018. **14**(6): p. 345-355.
189. Ben Maamar, M., et al., *Alterations in sperm DNA methylation, non-coding RNA expression, and histone retention mediate vinclozolin-induced epigenetic transgenerational inheritance of disease.* Environ Epigenet, 2018. **4**(2): p. dvy010.
190. Sharma, U. and O.J. Rando, *Metabolic Inputs into the Epigenome.* Cell Metab, 2017. **25**(3): p. 544-558.
191. Etchegaray, J.P. and R. Mostoslavsky, *Interplay between Metabolism and Epigenetics: A Nuclear Adaptation to Environmental Changes.* Mol Cell, 2016. **62**(5): p. 695-711.
192. Kaelin, W.G., Jr. and S.L. McKnight, *Influence of metabolism on epigenetics and disease.* Cell, 2013. **153**(1): p. 56-69.
193. Wolff, G.L., et al., *Maternal epigenetics and methyl supplements affect agouti gene expression in Avy/a mice.* FASEB J, 1998. **12**(11): p. 949-57.
194. Cooney, C.A., A.A. Dave, and G.L. Wolff, *Maternal methyl supplements in mice affect epigenetic variation and DNA methylation of offspring.* J Nutr, 2002. **132**(8 Suppl): p. 2393S-2400S.
195. Waterland, R.A., et al., *Maternal methyl supplements increase offspring DNA methylation at Axin Fused.* Genesis, 2006. **44**(9): p. 401-6.
196. Pauwels, S., et al., *The effect of paternal methyl-group donor intake on offspring DNA methylation and birth weight.* J Dev Orig Health Dis, 2017. **8**(3): p. 311-321.
197. Pauwels, S., et al., *Maternal intake of methyl-group donors affects DNA methylation of metabolic genes in infants.* Clin Epigenetics, 2017. **9**: p. 16.
198. Joubert, B.R., et al., *Maternal plasma folate impacts differential DNA methylation in an epigenome-wide meta-analysis of newborns.* Nat Commun, 2016. **7**: p. 10577.
199. Chamberlain, J.A., et al., *Dietary intake of one-carbon metabolism nutrients and DNA methylation in peripheral blood.* Am J Clin Nutr, 2018. **108**(3): p. 611-621.
200. Yan, H., et al., *IDH1 and IDH2 mutations in gliomas.* N Engl J Med, 2009. **360**(8): p. 765-73.
201. Dang, L., et al., *Cancer-associated IDH1 mutations produce 2-hydroxyglutarate.* Nature, 2009. **462**(7274): p. 739-44.
202. Figueroa, M.E., et al., *Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation.* Cancer Cell, 2010. **18**(6): p. 553-67.

203. Xiao, M., et al., *Inhibition of alpha-KG-dependent histone and DNA demethylases by fumarate and succinate that are accumulated in mutations of FH and SDH tumor suppressors.* Genes Dev, 2012. **26**(12): p. 1326-38.
204. Young, J.I., S. Zuchner, and G. Wang, *Regulation of the Epigenome by Vitamin C.* Annu Rev Nutr, 2015. **35**: p. 545-64.
205. Thienpont, B., et al., *Tumour hypoxia causes DNA hypermethylation by reducing TET activity.* Nature, 2016. **537**(7618): p. 63-68.
206. Mariani, C.J., et al., *TET1-mediated hydroxymethylation facilitates hypoxic gene induction in neuroblastoma.* Cell Rep, 2014. **7**(5): p. 1343-1352.
207. Niu, Y., et al., *Oxidative stress alters global histone modification and DNA methylation.* Free Radic Biol Med, 2015. **82**: p. 22-8.
208. Delatte, B., et al., *Genome-wide hydroxymethylcytosine pattern changes in response to oxidative stress.* Sci Rep, 2015. **5**: p. 12714.
209. Szarc vel Szic, K., et al., *From inflammaging to healthy aging by dietary lifestyle choices: is epigenetics the key to personalized nutrition?* Clin Epigenetics, 2015. **7**: p. 33.
210. Fang, M.Z., et al., *Tea polyphenol (-)-epigallocatechin-3-gallate inhibits DNA methyltransferase and reactivates methylation-silenced genes in cancer cell lines.* Cancer Res, 2003. **63**(22): p. 7563-70.
211. Zhu, H., G. Wang, and J. Qian, *Transcription factors as readers and effectors of DNA methylation.* Nat Rev Genet, 2016. **17**(9): p. 551-65.
212. Lappalainen, T. and J.M. Greally, *Associating cellular epigenetic models with human phenotypes.* Nat Rev Genet, 2017. **18**(7): p. 441-451.
213. Martin, E.M. and R.C. Fry, *A cross-study analysis of prenatal exposures to environmental contaminants and the epigenome: support for stress-responsive transcription factor occupancy as a mediator of gene-specific CpG methylation patterning.* Environ Epigenet, 2016. **2**(1).
214. Wan, J., et al., *Characterization of tissue-specific differential DNA methylation suggests distinct modes of positive and negative gene expression regulation.* BMC Genomics, 2015. **16**: p. 49.
215. Weber, M., et al., *Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome.* Nat Genet, 2007. **39**(4): p. 457-66.
216. Jones, P.A., *Functions of DNA methylation: islands, start sites, gene bodies and beyond.* Nat Rev Genet, 2012. **13**(7): p. 484-92.
217. Shen, L., et al., *Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters.* PLoS Genet, 2007. **3**(10): p. 2023-36.
218. Illingworth, R.S., et al., *Orphan CpG islands identify numerous conserved promoters in the mammalian genome.* PLoS Genet, 2010. **6**(9): p. e1001134.
219. Strathdee, G., et al., *Cell type-specific methylation of an intronic CpG island controls expression of the MCJ gene.* Carcinogenesis, 2004. **25**(5): p. 693-701.
220. Lee, S.M., et al., *Intragenic CpG islands play important roles in bivalent chromatin assembly of developmental genes.* Proc Natl Acad Sci U S A, 2017. **114**(10): p. E1885-E1894.
221. Neri, F., et al., *Intragenic DNA methylation prevents spurious transcription initiation.* Nature, 2017. **543**(7643): p. 72-77.
222. Jeziorska, D.M., et al., *DNA methylation of intragenic CpG islands depends on their transcriptional activity during differentiation and disease.* Proc Natl Acad Sci U S A, 2017. **114**(36): p. E7526-E7535.
223. Sarda, S., et al., *Distal CpG islands can serve as alternative promoters to transcribe genes with silenced proximal promoters.* Genome Res, 2017. **27**(4): p. 553-566.
224. Han, H., et al., *DNA methylation directly silences genes with non-CpG island promoters and establishes a nucleosome occupied promoter.* Hum Mol Genet, 2011. **20**(22): p. 4299-310.
225. Vire, E., et al., *The Polycomb group protein EZH2 directly controls DNA methylation.* Nature, 2006. **439**(7078): p. 871-4.
226. Mohn, F., et al., *Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors.* Mol Cell, 2008. **30**(6): p. 755-66.
227. Lock, L.F., N. Takagi, and G.R. Martin, *Methylation of the Hprt gene on the inactive X occurs after chromosome inactivation.* Cell, 1987. **48**(1): p. 39-46.

228. Charlet, J., et al., *Bivalent Regions of Cytosine Methylation and H3K27 Acetylation Suggest an Active Role for DNA Methylation at Enhancers.* Mol Cell, 2016. **62**(3): p. 422-431.
229. King, A.D., et al., *Reversible Regulation of Promoter and Enhancer Histone Landscape by DNA Methylation in Mouse Embryonic Stem Cells.* Cell Rep, 2016. **17**(1): p. 289-302.
230. Du, Q., et al., *Methyl-CpG-binding domain proteins: readers of the epigenome.* Epigenomics, 2015. **7**(6): p. 1051-73.
231. Baubec, T., et al., *Methylation-dependent and -independent genomic targeting principles of the MBD protein family.* Cell, 2013. **153**(2): p. 480-92.
232. Maurano, M.T., et al., *Role of DNA Methylation in Modulating Transcription Factor Occupancy.* Cell Rep, 2015. **12**(7): p. 1184-95.
233. Baralle, F.E. and J. Giudice, *Alternative splicing as a regulator of development and tissue identity.* Nat Rev Mol Cell Biol, 2017. **18**(7): p. 437-451.
234. Lev Maor, G., A. Yearim, and G. Ast, *The alternative role of DNA methylation in splicing regulation.* Trends Genet, 2015. **31**(5): p. 274-80.
235. Hodges, E., et al., *High definition profiling of mammalian DNA methylation by array capture and single molecule bisulfite sequencing.* Genome Res, 2009. **19**(9): p. 1593-605.
236. Yearim, A., et al., *HP1 is involved in regulating the global impact of DNA methylation on alternative splicing.* Cell Rep, 2015. **10**(7): p. 1122-34.
237. Ding, X.L., et al., *Isoform switching and exon skipping induced by the DNA methylation inhibitor 5-Aza-2'-deoxycytidine.* Sci Rep, 2016. **6**: p. 24545.
238. Shayevitch, R., et al., *The importance of DNA methylation of exons on alternative splicing.* RNA, 2018. **24**(10): p. 1351-1362.
239. Maunakea, A.K., et al., *Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition.* Cell Res, 2013. **23**(11): p. 1256-69.
240. Shukla, S., et al., *CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing.* Nature, 2011. **479**(7371): p. 74-9.
241. Wong, J.J., et al., *Intron retention is regulated by altered MeCP2-mediated splicing factor recruitment.* Nat Commun, 2017. **8**: p. 15134.
242. Ferguson-Smith, A.C., *Genomic imprinting: the emergence of an epigenetic paradigm.* Nat Rev Genet, 2011. **12**(8): p. 565-75.
243. Peters, J., *The role of genomic imprinting in biology and disease: an expanding view.* Nat Rev Genet, 2014. **15**(8): p. 517-30.
244. Morison, I.M., C.J. Paton, and S.D. Cleverley, *The imprinted gene and parent-of-origin effect database.* Nucleic Acids Res, 2001. **29**(1): p. 275-6.
245. Baran, Y., et al., *The landscape of genomic imprinting across diverse adult human tissues.* Genome Res, 2015. **25**(7): p. 927-36.
246. Travers, M.E., et al., *Insights into the molecular mechanism for type 2 diabetes susceptibility at the KCNQ1 locus from temporal changes in imprinting status in human islets.* Diabetes, 2013. **62**(3): p. 987-92.
247. Haig, D., *Coadaptation and conflict, misconception and muddle, in the evolution of genomic imprinting.* Heredity (Edinb), 2014. **113**(2): p. 96-103.
248. Wolf, J.B. and R. Hager, *A maternal-offspring coadaptation theory for the evolution of genomic imprinting.* PLoS Biol, 2006. **4**(12): p. e380.
249. Moore, T. and D. Haig, *Genomic imprinting in mammalian development: a parental tug-of-war.* Trends Genet, 1991. **7**(2): p. 45-9.
250. Iwasa, Y., *The conflict theory of genomic imprinting: how much can be explained?* Curr Top Dev Biol, 1998. **40**: p. 255-93.
251. Wilkins, J.F. and D. Haig, *What good is genomic imprinting: the function of parent-specific gene expression.* Nat Rev Genet, 2003. **4**(5): p. 359-68.
252. Barlow, D.P. and M.S. Bartolomei, *Genomic imprinting in mammals.* Cold Spring Harb Perspect Biol, 2014. **6**(2).
253. Chotalia, M., et al., *Transcription is required for establishment of germline methylation marks at imprinted genes.* Genes Dev, 2009. **23**(1): p. 105-17.

254. Henckel, A., et al., *Transcription and histone methylation changes correlate with imprint acquisition in male germ cells.* EMBO J, 2012. **31**(3): p. 606-15.
255. Quenneville, S., et al., *In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions.* Mol Cell, 2011. **44**(3): p. 361-72.
256. Li, X., et al., *A maternal-zygotic effect gene, Zfp57, maintains both maternal and paternal imprints.* Dev Cell, 2008. **15**(4): p. 547-57.
257. Nakamura, T., et al., *PGC7/Stella protects against DNA demethylation in early embryogenesis.* Nat Cell Biol, 2007. **9**(1): p. 64-71.
258. Zhang, T., et al., *G9a/GLP Complex Maintains Imprinted DNA Methylation in Embryonic Stem Cells.* Cell Rep, 2016. **15**(1): p. 77-85.
259. Mortier Geert, V.B.W., *Genomics, epigenetics and growth*, in *Human Growth and Development*, B.B. Noël Cameron, Editor. 2012.
260. Walsh, C.P., J.R. Chaillet, and T.H. Bestor, *Transcription of IAP endogenous retroviruses is constrained by cytosine methylation.* Nat Genet, 1998. **20**(2): p. 116-7.
261. Bourc'his, D. and T.H. Bestor, *Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L.* Nature, 2004. **431**(7004): p. 96-9.
262. Barau, J., et al., *The DNA methyltransferase DNMT3C protects male germ cells from transposon activity.* Science, 2016. **354**(6314): p. 909-912.
263. Aravin, A.A., et al., *A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice.* Mol Cell, 2008. **31**(6): p. 785-99.
264. Aravin, A.A., et al., *Developmentally regulated piRNA clusters implicate MILI in transposon control.* Science, 2007. **316**(5825): p. 744-7.
265. Kuramochi-Miyagawa, S., et al., *DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes.* Genes Dev, 2008. **22**(7): p. 908-17.
266. Ozata, D.M., et al., *PIWI-interacting RNAs: small RNAs with big functions.* Nat Rev Genet, 2019. **20**(2): p. 89-108.
267. Bakshi, A., et al., *DNA methylation variation of human-specific Alu repeats.* Epigenetics, 2016. **11**(2): p. 163-73.
268. Varshney, D., et al., *SINE transcription by RNA polymerase III is suppressed by histone methylation but not by DNA methylation.* Nat Commun, 2015. **6**: p. 6569.
269. Eden, A., et al., *Chromosomal instability and tumors promoted by DNA hypomethylation.* Science, 2003. **300**(5618): p. 455.
270. Gaudet, F., et al., *Induction of tumors in mice by genomic hypomethylation.* Science, 2003. **300**(5618): p. 489-92.
271. Chen, R.Z., et al., *DNA hypomethylation leads to elevated mutation rates.* Nature, 1998. **395**(6697): p. 89-93.
272. Karpf, A.R. and S. Matsui, *Genetic disruption of cytosine DNA methyltransferase enzymes induces chromosomal instability in human cancer cells.* Cancer Res, 2005. **65**(19): p. 8635-9.
273. Xu, G.L., et al., *Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene.* Nature, 1999. **402**(6758): p. 187-91.
274. Gonzalo, S., et al., *DNA methyltransferases control telomere length and telomere recombination in mammalian cells.* Nat Cell Biol, 2006. **8**(4): p. 416-24.
275. Jin, B. and K.D. Robertson, *DNA methyltransferases, DNA damage repair, and cancer.* Adv Exp Med Biol, 2013. **754**: p. 3-29.
276. Mortusewicz, O., et al., *Recruitment of DNA methyltransferase I to DNA repair sites.* Proc Natl Acad Sci U S A, 2005. **102**(25): p. 8905-9.
277. Ha, K., et al., *Rapid and transient recruitment of DNMT1 to DNA double-strand breaks is mediated by its interaction with multiple components of the DNA damage response machinery.* Hum Mol Genet, 2011. **20**(1): p. 126-40.
278. Kim, M., et al., *Dnmt1 deficiency leads to enhanced microsatellite instability in mouse embryonic stem cells.* Nucleic Acids Res, 2004. **32**(19): p. 5742-9.
279. Guo, G., W. Wang, and A. Bradley, *Mismatch repair genes identified using genetic screens in Blm-deficient embryonic stem cells.* Nature, 2004. **429**(6994): p. 891-5.

280. Loughery, J.E., et al., *DNMT1 deficiency triggers mismatch repair defects in human cells through depletion of repair protein levels in a process involving the DNA damage response.* Hum Mol Genet, 2011. **20**(16): p. 3241-55.
281. Kafer, G.R., et al., *5-Hydroxymethylcytosine Marks Sites of DNA Damage and Promotes Genome Stability.* Cell Rep, 2016. **14**(6): p. 1283-1292.
282. Jiang, D., et al., *TET3-mediated DNA oxidation promotes ATR-dependent DNA damage response.* EMBO Rep, 2017. **18**(5): p. 781-796.
283. Coulter, J.B., et al., *TET1 deficiency attenuates the DNA damage response and promotes resistance to DNA damaging agents.* Epigenetics, 2017. **12**(10): p. 854-864.
284. You, J.S. and P.A. Jones, *Cancer genetics and epigenetics: two sides of the same coin?* Cancer Cell, 2012. **22**(1): p. 9-20.
285. Yang, L., et al., *DNMT3A Loss Drives Enhancer Hypomethylation in FLT3-ITD-Associated Leukemias.* Cancer Cell, 2016. **29**(6): p. 922-934.
286. Yamazaki, J., et al., *TET2 Mutations Affect Non-CpG Island DNA Methylation at Enhancers and Transcription Factor-Binding Sites in Chronic Myelomonocytic Leukemia.* Cancer Res, 2015. **75**(14): p. 2833-43.
287. Feinberg, A.P., *The Key Role of Epigenetics in Human Disease Prevention and Mitigation.* N Engl J Med, 2018. **378**(14): p. 1323-1334.
288. Klutstein, M., et al., *DNA Methylation in Cancer and Aging.* Cancer Res, 2016. **76**(12): p. 3446-50.
289. Keshet, I., et al., *Evidence for an instructive mechanism of de novo methylation in cancer cells.* Nat Genet, 2006. **38**(2): p. 149-53.
290. Schlesinger, Y., et al., *Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer.* Nat Genet, 2007. **39**(2): p. 232-6.
291. Widschwendter, M., et al., *Epigenetic stem cell signature in cancer.* Nat Genet, 2007. **39**(2): p. 157-8.
292. Vidal, E., et al., *A DNA methylation map of human cancer at single base-pair resolution.* Oncogene, 2017. **36**(40): p. 5648-5657.
293. Heyn, H., et al., *Epigenomic analysis detects aberrant super-enhancer DNA methylation in human cancer.* Genome Biol, 2016. **17**: p. 11.
294. Agirre, X., et al., *Whole-epigenome analysis in multiple myeloma reveals DNA hypermethylation of B cell-specific enhancers.* Genome Res, 2015. **25**(4): p. 478-87.
295. Wen, B., et al., *Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells.* Nat Genet, 2009. **41**(2): p. 246-50.
296. Zhou, W., et al., *DNA methylation loss in late-replicating domains is linked to mitotic cell division.* Nat Genet, 2018. **50**(4): p. 591-602.
297. Berman, B.P., et al., *Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains.* Nat Genet, 2011. **44**(1): p. 40-6.
298. Timp, W., et al., *Large hypomethylated blocks as a universal defining epigenetic alteration in human solid tumors.* Genome Med, 2014. **6**(8): p. 61.
299. Donaldson-Collier, M.C., et al., *EZH2 oncogenic mutations drive epigenetic, transcriptional, and structural changes within chromatin domains.* Nat Genet, 2019. **51**(3): p. 517-528.
300. De, S. and F. Michor, *DNA secondary structures and epigenetic determinants of cancer genome evolution.* Nat Struct Mol Biol, 2011. **18**(8): p. 950-5.
301. Daskalos, A., et al., *Hypomethylation of retrotransposable elements correlates with genomic instability in non-small cell lung cancer.* Int J Cancer, 2009. **124**(1): p. 81-7.
302. Howard, G., et al., *Activation and transposition of endogenous retroviral elements in hypomethylation induced tumors in mice.* Oncogene, 2008. **27**(3): p. 404-8.
303. Van Tongelen, A., A. Loriot, and C. De Smet, *Oncogenic roles of DNA hypomethylation through the activation of cancer-germline genes.* Cancer Lett, 2017. **396**: p. 130-137.
304. Flavahan, W.A., E. Gaskell, and B.E. Bernstein, *Epigenetic plasticity and the hallmarks of cancer.* Science, 2017. **357**(6348).
305. Pfeifer, G.P., *Defining Driver DNA Methylation Changes in Human Cancer.* Int J Mol Sci, 2018. **19**(4).

306. Timp, W. and A.P. Feinberg, *Cancer as a dysregulated epigenome allowing cellular growth advantage at the expense of the host.* Nat Rev Cancer, 2013. **13**(7): p. 497-510.
307. Landau, D.A., et al., *Locally disordered methylation forms the basis of intratumor methylome variation in chronic lymphocytic leukemia.* Cancer Cell, 2014. **26**(6): p. 813-825.
308. Hansen, K.D., et al., *Increased methylation variation in epigenetic domains across cancer types.* Nat Genet, 2011. **43**(8): p. 768-75.
309. Teschendorff, A.E., et al., *Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer.* Genome Res, 2010. **20**(4): p. 440-6.
310. Yuan, T., et al., *An integrative multi-scale analysis of the dynamic DNA methylation landscape in aging.* PLoS Genet, 2015. **11**(2): p. e1004996.
311. Vandiver, A.R., et al., *Age and sun exposure-related widespread genomic blocks of hypomethylation in nonmalignant skin.* Genome Biol, 2015. **16**: p. 80.
312. Hanley, M.P., et al., *Genome-wide DNA methylation profiling reveals cancer-associated changes within early colonic neoplasia.* Oncogene, 2017. **36**(35): p. 5035-5044.
313. Teschendorff, A.E., et al., *Epigenetic variability in cells of normal cytology is associated with the risk of future morphological transformation.* Genome Med, 2012. **4**(3): p. 24.
314. Mazor, T., et al., *Intratumoral Heterogeneity of the Epigenome.* Cancer Cell, 2016. **29**(4): p. 440-451.
315. Alderton, G.K., *Tumour evolution: Epigenetic and genetic heterogeneity in metastasis.* Nat Rev Cancer, 2017. **17**(3): p. 141.
316. Michels, K.B., et al., *Recommendations for the design and analysis of epigenome-wide association studies.* Nat Methods, 2013. **10**(10): p. 949-55.
317. Ahuja, Y.R., S. Sharma, and V. Mohan, *Cardiovascular diseases: Interplay of epigenetics.* Clin Exp Hypertens, 2017. **39**(1): p. 1-7.
318. van Dijk, S.J., et al., *Epigenetics and human obesity.* Int J Obes (Lond), 2015. **39**(1): p. 85-97.
319. Kwak, S.H. and K.S. Park, *Recent progress in genetic and epigenetic research on type 2 diabetes.* Exp Mol Med, 2016. **48**: p. e220.
320. Qazi, T.J., et al., *Epigenetics in Alzheimer's Disease: Perspective of DNA Methylation.* Mol Neurobiol, 2018. **55**(2): p. 1026-1044.
321. Pavlou, M.A.S. and T.F. Outeiro, *Epigenetics in Parkinson's Disease.* Adv Exp Med Biol, 2017. **978**: p. 363-390.
322. Zheleznyakova, G.Y., et al., *Epigenetic research in multiple sclerosis: progress, challenges, and opportunities.* Physiol Genomics, 2017. **49**(9): p. 447-461.
323. DeVries, A. and D. Vercelli, *Epigenetic Mechanisms in Asthma.* Ann Am Thorac Soc, 2016. **13 Suppl 1**: p. S48-50.
324. Leenen, F.A., C.P. Muller, and J.D. Turner, *DNA methylation: conducting the orchestra from exposure to phenotype?* Clin Epigenetics, 2016. **8**: p. 92.
325. Relton, C.L. and G. Davey Smith, *Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease.* Int J Epidemiol, 2012. **41**(1): p. 161-76.
326. Wahl, S., et al., *Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity.* Nature, 2017. **541**(7635): p. 81-86.
327. Teschendorff, A.E. and S.C. Zheng, *Cell-type deconvolution in epigenome-wide association studies: a review and recommendations.* Epigenomics, 2017. **9**(5): p. 757-768.
328. consortium, B., *Quantitative comparison of DNA methylation assays for biomarker development and clinical applications.* Nat Biotechnol, 2016. **34**(7): p. 726-37.
329. Zeng, H., et al., *Liquid biopsies: DNA methylation analyses in circulating cell-free DNA.* J Genet Genomics, 2018. **45**(4): p. 185-192.
330. Tanic, M. and S. Beck, *Epigenome-wide association studies for cancer biomarker discovery in circulating cell-free DNA: technical advances and challenges.* Curr Opin Genet Dev, 2017. **42**: p. 48-55.
331. Lehmann-Werman, R., et al., *Identification of tissue-specific cell death using methylation patterns of circulating DNA.* Proc Natl Acad Sci U S A, 2016. **113**(13): p. E1826-34.

332. Berdasco, M. and M. Esteller, *Clinical epigenetics: seizing opportunities for translation.* Nat Rev Genet, 2019. **20**(2): p. 109-127.
333. Ned, R.M., S. Melillo, and M. Marrone, *Fecal DNA testing for Colorectal Cancer Screening: the ColoSure test.* PLoS Curr, 2011. **3**: p. RRN1220.
334. Imperiale, T.F., D.F. Ransohoff, and S.H. Itzkowitz, *Multitarget stool DNA testing for colorectal-cancer screening.* N Engl J Med, 2014. **371**(2): p. 187-8.
335. Warren, J.D., et al., *Septin 9 methylated DNA is a sensitive and specific blood test for colorectal cancer.* BMC Med, 2011. **9**: p. 133.
336. Lamb, Y.N. and S. Dhillon, *Epi proColon((R)) 2.0 CE: A Blood-Based Screening Test for Colorectal Cancer.* Mol Diagn Ther, 2017. **21**(2): p. 225-232.
337. Hegi, M.E., et al., *MGMT gene silencing and benefit from temozolomide in glioblastoma.* N Engl J Med, 2005. **352**(10): p. 997-1003.
338. Esteller, M., et al., *Inactivation of the DNA-repair gene MGMT and the clinical response of gliomas to alkylating agents.* N Engl J Med, 2000. **343**(19): p. 1350-4.
339. Moran, S., et al., *Epigenetic profiling to classify cancer of unknown primary: a multicentre, retrospective analysis.* Lancet Oncol, 2016. **17**(10): p. 1386-1395.
340. Lunnon, K., et al., *Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease.* Nat Neurosci, 2014. **17**(9): p. 1164-70.
341. de Mello, V.D., et al., *DNA methylation in obesity and type 2 diabetes.* Ann Med, 2014. **46**(3): p. 103-13.
342. Ronn, T., et al., *A six months exercise intervention influences the genome-wide DNA methylation pattern in human adipose tissue.* PLoS Genet, 2013. **9**(6): p. e1003572.
343. Horvath, S. and K. Raj, *DNA methylation-based biomarkers and the epigenetic clock theory of ageing.* Nat Rev Genet, 2018. **19**(6): p. 371-384.
344. Quach, A., et al., *Epigenetic clock analysis of diet, exercise, education, and lifestyle factors.* Aging (Albany NY), 2017. **9**(2): p. 419-446.
345. Maegawa, S., et al., *Caloric restriction delays age-related methylation drift.* Nat Commun, 2017. **8**(1): p. 539.
346. Zannas, A.S., et al., *Lifetime stress accelerates epigenetic aging in an urban, African American cohort: relevance of glucocorticoid signaling.* Genome Biol, 2015. **16**: p. 266.
347. Gale, C.R., et al., *The epigenetic clock and objectively measured sedentary and walking behavior in older adults: the Lothian Birth Cohort 1936.* Clin Epigenetics, 2018. **10**: p. 4.
348. Diesch, J., et al., *A clinical-molecular update on azanucleoside-based therapy for the treatment of hematologic cancers.* Clin Epigenetics, 2016. **8**: p. 71.
349. Gnyszka, A., Z. Jastrzebski, and S. Flis, *DNA methyltransferase inhibitors and their emerging role in epigenetic therapy of cancer.* Anticancer Res, 2013. **33**(8): p. 2989-96.
350. Falkenberg, K.J. and R.W. Johnstone, *Histone deacetylases and their inhibitors in cancer, neurological diseases and immune disorders.* Nat Rev Drug Discov, 2014. **13**(9): p. 673-91.
351. Delmore, J.E., et al., *BET bromodomain inhibition as a therapeutic strategy to target c-Myc.* Cell, 2011. **146**(6): p. 904-17.
352. McCabe, M.T., et al., *EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations.* Nature, 2012. **492**(7427): p. 108-12.
353. Stein, E.M., *IDH2 inhibition in AML: Finally progress?* Best Pract Res Clin Haematol, 2015. **28**(2-3): p. 112-5.
354. Daigle, S.R., et al., *Selective killing of mixed lineage leukemia cells by a potent small-molecule DOT1L inhibitor.* Cancer Cell, 2011. **20**(1): p. 53-65.
355. Mohammad, H.P., et al., *A DNA Hypomethylation Signature Predicts Antitumor Activity of LSD1 Inhibitors in SCLC.* Cancer Cell, 2015. **28**(1): p. 57-69.
356. Brocks, D., et al., *DNMT and HDAC inhibitors induce cryptic transcription start sites encoded in long terminal repeats.* Nat Genet, 2017. **49**(7): p. 1052-1060.
357. Jones, P.A., et al., *Epigenetic therapy in immune-oncology.* Nat Rev Cancer, 2019. **19**(3): p. 151-161.
358. Brocken, D.J.W., M. Tark-Dame, and R.T. Dame, *dCas9: A Versatile Tool for Epigenome Editing.* Curr Issues Mol Biol, 2018. **26**: p. 15-32.

2

Thesis outline and objectives

As outlined in the introductory section, DNA methylation is thought to mediate the long-term effects of chronic exposure to environmental factors on health and disease. Due to its dynamic nature, DNA methylation can be exploited as prognostic, diagnostic and therapeutic biomarkers. Besides, reversible DNA methylation changes are also an attractive target for therapeutic and nutritional interventions. In addition, studying epigenetic patterns could give us useful mechanistic insight in disease etiology and development. In this thesis, we studied DNA methylation patterns in easy-accessible tissues, including blood and saliva to monitor adverse early environmental exposures, cardio-metabolic diseases and nutritional intervention responses. The results section of the thesis aims to answer 3 research questions-objectives:

PART 1: Can we monitor early adverse environmental conditions using DNA methylation marks in blood and/or saliva?

During prenatal and early development, the methylation profile is sensitive to adverse environmental factors and the exposure history builds up cellular memory which can promote diseases later in life. In this way, monitoring adverse DNA methylation changes holds promise to monitor and/or prevent potential disease risk later in life. In the first part of the result section, we examine whether early adverse environmental conditions can be monitored through DNA methylation marks in blood and saliva.

In **chapter 3**, we performed a pilot study in a cohort of Danish children, whose mothers were exposed to pesticides during pregnancy. In a previous study of those children, it was found that carriers of a polymorphism in the PON1 enzyme, which has an important role in organophosphate pesticide hydrolysis and is linked with atherosclerosis, have an adverse cardio-metabolic risk profile upon pesticide exposure. We explored whether we could find a specific DNA methylation profile associated with prenatal pesticide exposure and PON1 polymorphism. Furthermore we evaluated whether we can link this profile with the cardio-metabolic risk factors observed in these children.

In **chapter 4**, we compared DNA methylation profiles in saliva samples of neglected orphan children of the well characterized Bucharest cohort, raised in institutions, following foster care or never-institutionalized children. We addressed whether we could find a salivary DNA methylation profile associated with institutionalization associated stress and evaluated to which extent foster care can remediate the stress related epigenetic effects observed in institutionalized children. Finally epigenetic changes were cross-compared with variations in gene expression and cortisol levels to determine significant associations.

PART 2: Can we identify epigenetic blood surrogate markers to predict cardio-metabolic diseases?

Epigenetics is believed to play a significant role in the initiation and development of complex chronic diseases. For some diseases, the tissue of interest is difficult or even impossible to obtain, and therefore easily accessible surrogate tissues, like blood and saliva are recommended. In the second part of the result section, we evaluated whether we could identify disease associated DNA methylation profiles in whole blood of atherosclerosis patients that could be used as a surrogate biomarker.

In **chapter 5**, we measured the genome-wide DNA methylation profile of whole blood samples of clinical atherosclerosis patients and compared it with epigenetic profiles of atherosclerotic plaque material, and subclinical atherosclerosis samples of the Aragon workers cohort, to identify potential surrogate markers for early CVD detection.

In **chapter 6**, we cross-compared the atherosclerosis specific methylation profile identified in chapter 5 with adverse DNA methylation profiles of other lifestyle diseases, to identify unique and/or general biomarkers for inflammaging type disorders in blood and/or solid tissues.

PART 3: Can we reverse adverse DNA methylation patterns by nutritional interventions?

Various environmental and lifestyle factors trigger cumulative beneficial and/or harmful epigenetic changes during life. We characterized whether adverse methylation changes could be reversed by changing dietary lifestyle. Diet and nutrition are important lifestyle factors which can have a significant impact on health and disease risk. In the last part of the result section, we studied the epigenetic effects of nutritional and immunomodulatory phytochemical compounds on DNA methylation. Moreover, we addressed to which extent disease associated DNA methylation patterns can be reversed by nutraceutical and/or phytomedicinal specific interventions.

In **chapter 7**, we examined DNA methylation changes upon *in vitro* exposure of endothelial cells to cardio-protective flavanols or in blood samples following an eight-week diet intervention with flavanol-rich grape seed extract. Furthermore, we searched for reversible CVD associated epigenetic changes in response to flavanols.

In **chapter 8**, we studied epigenetic changes upon *in vitro* exposure of THP1 monocytes to Echinaforce®, a commercial standardized immunomodulatory extract, which can protect against colds and respiratory infections. Finally, epigenetic changes were integrated with variations in gene expression and kinase signaling to resolve a mechanism of action.

3

Prenatal pesticide exposure interacts with a common polymorphism in the *PON1* gene leading to specific DNA methylation changes in genes associated with cardio-metabolic disease risk

Ken Declerck[1], Sylvie Remy[2,3], Christine Wohlfahrt-Veje[4], Katharina M. Main[4], Guy Van Camp[5], Greet Schoeters[3,6,7], Wim Vanden Berghe[1], Helle R. Andersen[7]

[1] Laboratory of Protein Chemistry, Proteomics and Epigenetic Signalling (PPES), Department of Biomedical Sciences, University of Antwerp, Universiteitsplein 1, Antwerp, Belgium
[2] Department of Epidemiology and Social Medicine, Antwerp University, Universiteitsplein 1, Antwerp, Belgium
[3] Flemish Institute for Technological Research (VITO), Unit Environmental Risk and Health, Boeretang 200, Mol, Belgium
[4] University Department of Growth and Reproduction, University Hospital of Copenhagen, Rigshospitalet, Copenhagen, Denmark
[5] Center of Medical Genetics, University of Antwerp & Antwerp University Hospital, Antwerp, Belgium
[6] Department of Biomedical Sciences, Antwerp University, Universiteitsplein 1, Antwerp, Belgium
[7] Environmental Medicine, Institute of Public Health, University of Southern Denmark, Odense, Denmark

Published in Clinical epigenetics (https://doi.org/10.1186/s13148-017-0336-4)

Contribution: Ken Declerck performed the experiments, analysed the data, helped in interpreting the results and wrote the manuscript.

Abstract | Prenatal environmental conditions may influence disease risk in later life. Previously, a gene-environment interaction was demonstrated between the paraoxonase 1 (*PON1)* Q192R genotype and prenatal pesticide exposure, which increases adverse cardio-metabolic risk profile at school age. However, the molecular mechanisms involved have not yet been resolved. It was hypothesized that epigenetics might be involved. The aim of the present study was therefore to investigate whether DNA methylation patterns in blood cells were related to prenatal pesticide exposure level, *PON1* Q192R genotype, and associated metabolic effects observed in the children.

Whole blood DNA methylation patterns in 48 children (6-11 years of age), whose mothers were occupationally unexposed or exposed to pesticides early in pregnancy, were determined by Illumina 450K methylation arrays.

A specific methylation profile was observed in prenatally pesticide exposed children carrying the *PON1* 192R-allele. Differentially methylated genes were enriched in several neuroendocrine signaling pathways including dopamine-DARPP32 feedback (appetite, reward pathways), corticotrophin releasing hormone signalling, nNOS, neuregulin signalling, mTOR signalling and type II diabetes mellitus signalling suggesting a possible link with the metabolic effects observed in these children. Furthermore, we were able to identify possible candidate genes which mediated the associations between pesticide exposure and increased leptin level, body fat percentage, and difference in BMI Z-score between birth and school age.

DNA methylation may be an underlying mechanism explaining an adverse cardio-metabolic health profile in children carrying the *PON1* 192R-allele and prenatally exposed to pesticides.

Introduction

A considerable part of modern pesticides has neurotoxic and/or endocrine disrupting properties [1-3] and therefore the potential to disturb development of neurobehavioral, neuroendocrine, and reproductive functions [4-8] especially if exposure occurs during vulnerable time periods in fetal life or early childhood. To investigate potential health effects of prenatal pesticide exposure, we have followed a cohort of children, whose mothers were employed in greenhouse horticulture in pregnancy. Some of the mothers were occupationally exposed to mixtures of pesticides in the first trimester before the pregnancy was recognized and preventive measures were taken. Findings from this cohort include associations between maternal pesticide exposure and lower birth weight followed by increased body fat accumulation during

childhood [9], impaired reproductive development in boys [10, 11], and earlier breast development [12] and impaired neurobehavioral function in girls [13].

The HDL-associated enzyme paraoxonase 1 (PON1) catalyzes the hydrolysis of a wide range of substrates including some organophosphate insecticides [14, 15]. It also protects lipoproteins from oxidative modifications and hence against development of atherosclerosis [16, 17]. A common polymorphism in the coding sequence of the *PON1* gene substitutes glutamine (Q) to arginine (R) at position 192. This substitution seems to affect both properties of the enzyme and several studies have indicated an increased risk of CVD in R-allele carriers [17, 18]. To investigate if this polymorphism affected the sensitivity to prenatal pesticide exposure, the *PON1* Q192R genotype was determined in the children. We found a marked interaction between prenatal pesticide exposure and the *PON1* Q192R genotype. At school age, exposed children with the R-allele had significantly higher BMI, body fat percentage, abdominal circumference, and blood pressure compared to unexposed children with the same genotype. In the group of children with the QQ genotype, there was no effect of prenatal pesticide exposure on these parameters [19]. In addition, serum concentrations of leptin, glucagon, and plasminogen activator inhibitor type-1 (PAI-1) were enhanced in prenatally pesticide exposed children with the R-allele, also after adjusting for BMI [20] which also indicates disturbance of metabolic pathways related to development of metabolic syndrome [21-23]. In addition, leptin seemed to be a mediator of the increased fat accumulation during childhood related to prenatal pesticide exposure in children with the *PON1* 192R-allele [20]. Thus, the obtained results indicate a gene-environment interaction between pesticide exposure and *PON1* gene heterogeneities already in early prenatal life that might enhance the risk of cardio-metabolic diseases later in life.

The mechanism behind this interaction is not yet understood but might be mediated by epigenetic alterations depending on both genotype and prenatal exposure. Epigenetic marks, including DNA methylation and covalent histone modifications, are dynamic and can adapt to a variety of external stimuli [24]. Furthermore, during fetal development extensive de- and re-methylation events are taking place making this period highly vulnerable for epigenetic changes caused by environmental conditions [25]. Indeed, emerging evidence in experimental animals and in humans associate altered DNA methylation patterns with a variety of prenatal exposures including dietary factors, parental care, infections, smoking and environmental pollutants [26-31]. In experimental animals, early life changes in DNA methylation have been associated with diet induced obesity and insulin resistance [32]. Recently, also human studies have suggested that DNA methylation patterns at birth are related to birth weight and fat mass later in childhood [33, 34]. The aim of this exploratory study was to investigate whether methylation patterns in blood samples of school children were related to

prenatal pesticide exposure, *PON1* Q192R genotype, and adverse health outcomes already observed in the children. We hypothesized that the health effects associated with early prenatal pesticide exposure were related to differential epigenetic modifications in children with the QQ-genotype and children carrying the R-allele.

Materials and methods

Study population

This study is a part of an ongoing prospective study including 203 children born between 1996 and 2001 by female greenhouse workers. The children were examined for the first time at three months of age [11] and followed up at school age when 44 new age-matched controls were included [9] and the *PON1* genotype was determined for 141 children [19]. For this exploratory study, 48 pre-pubertal (Tanner Stage 1) children, whose mothers reported not to have smoked during pregnancy, were selected equally distributed between the *PON1* 192QQ and QR/RR genotype. The QR/RR genotype group consisted of 3 children with the RR genotype and 21 with the QR genotype. After excluding children of mothers who smoked in pregnancy, the number of unexposed controls within each genotype was low, 20 with the QQ genotype and 16 with the QR/RR genotype. DNA qualified for methylation analysis was only available for 11 and 12 of these children, respectively. For each genotype, we then used individual matching to select one exposed child of same sex and age for each of the controls. For the QQ-genotype, two exposed children were selected for each of two controls to obtain 24 children. Thus, in total we used data from 13 exposed and 11 unexposed children with the QQ genotype, and 12 exposed and 12 unexposed children with the QR/RR genotype (**Table 1**).

Recruitment, characteristics, exposure categorization, and clinical examinations of the children have previously been described in detail [9, 11, 19]. Briefly, we recruited pregnant women working in greenhouses and referred to the local Department of Occupational Health for risk assessment of their working conditions and guidance for safe work practices during pregnancy. Detailed information about working conditions inclusive pesticide use for the previous three months was obtained from maternal interview at enrollment (gestational weeks 4-10) and supplemented by telephone contact to the employers. For all women, re-entry activities (such as moving or packing potted plants or nipping cuttings) constituted their main work functions. Approximately 20% of the women reported having been directly involved in applying pesticides, mainly by irrigating fungicides or growth retardants. Only few (6%) of the women had applied

insecticides. The women were categorized as occupationally exposed if pesticides were applied in the working area more than once a month and the women handled treated plants within one week after treatment and/or the women were directly involved in applying pesticides. The women were categorized as occupationally unexposed if none of the above criteria was fulfilled. All exposure assessments and categorization of the mothers as pesticide exposed or unexposed were performed independently by two toxicologists before the first examination of the children. Women categorized as pesticide exposed, went on paid leave or were moved to work functions with less or no pesticide exposure shortly after enrollment. Hence, the exposure classification relates to the early weeks of the first trimester before study enrollment.

The exposure situation was complex since the use of specific pesticides varied with time and location, both within the same company and between companies, depending on the plant production and the type of pest to be controlled. Out of 124 different active pesticide ingredients used in the greenhouses were 59 insecticides (17 organophosphates, 12 pyrethroids, 9 carbamates, and 21 others), 40 fungicides, 11 growth regulators, and 14 herbicides. Some were used only in few greenhouses or in short periods, whereas others were used more often. Organophosphate insecticides were used to some extent in the working areas for 91% of the exposed mothers in the entire cohort, and for 24 out of the 25 exposed mothers whose children were included in this study. The most used organophosphates were dichlorvos, dimethoate, and chlorpyrifos. Other frequently used pesticides were the pyrethroid insecticides deltamethrin and fenpropathrin; the carbamate insecticides methiocarb, pirimicarb, and methomyl, and the fungicides fenarimol, prochloraz, tolclofos-methyl, vinclozolin, iprodion, and chlorothalonil. In general, the time interval between applying insecticides and working in the treated areas was longer (1-3 days) than for fungicides and growth regulators (often a few hours). Because of the complexity of the exposure situation and because most of the women at enrollment had been off work for some days while the risk assessment of their working conditions was performed, biomonitoring of the exposure was not feasible. A complete list of the pesticides used in the greenhouses can be obtained from the corresponding author.

At follow-up at age 6 to 11 years, 177 children underwent a standardized clinical examination in which systolic and diastolic blood pressure, pubertal staging, height, weight, thickness of skin folds, and other anthropometric parameters were measured [9]. The same pediatrician performed all clinical examinations blinded to information about maternal pesticide exposure during pregnancy.

Venous non-fasting blood samples were collected (between midmorning and late afternoon) in EDTA coated and uncoated vials (Venoject). After centrifugation at 2000

g for 10 min at 20°C, buffy coat for genotyping and epigenetic analysis was separated from the EDTA-treated samples. Buffy coat and serum from the uncoated vials were stored at -80°C until analysis.
As previously described [19], C-108T (rs705379) and Q192R (rs662) polymorphisms of the *PON1* gene was determined by the Taqman-based allele discrimination using the ABI Prism 7700 Sequence Detection System, serum activity of *PON1* was determined by spectrophotometry with paraoxon as substrate, and insulin (proinsulin and insulin) and leptin concentrations in serum were determined by commercial ELISA hormone kits from RayBio.
Genotyping and all serum analyses were performed blinded to both exposure information and examination outcomes.

Ethics

The study was conducted according to the Helsinki II Declaration with written informed consent by all parents and oral consent by the children as approved by The Regional Scientific Ethical Committees for Southern Denmark (S-20070068) and the Danish Data Protection Agency.

Sample preparation

DNA from buffy coat samples was extracted using QIAamp DNA Blood mini kit (Qiagen, Hilden, Germany). The blood spin protocol was applied according to manufacturer's instructions. Samples were eluted in 100 µl elution buffer. DNA samples were bisulfite converted using the EZ DNA methylation kit from Zymo according to manufacturer's instructions. Successful bisulfite conversion was checked using a bisulfite-specific PCR of an amplicon in the *SALL3* gene (see **Supplementary Table 1** for primer sequences). Only samples showing an intense band on agarose gel were further analysed by the 450K methylation array. As a negative control non-converted gDNA was used.

DNA methylation and data preprocessing

The Infinium HumanMethylation450 BeadChip array (Illumina, San Diego, CA, USA) was used to measure DNA methylation genome-wide. 4 µL of bisulfite-converted DNA from each sample was amplified, fragmented, precipitated, resuspended and subsequently hybridized onto the BeadChips. After overnight incubation of the BeadChips, unhybridized fragments were washed away, while hybridized fragments were extended using fluorescent nucleotide bases. Finally, the BeadChips were scanned using the Illumina iScan system to obtain raw methylation intensities for each probe.

We used the R package RnBeads to preprocess the Illumina 450K methylation data [35]. Cg-probes were filtered before normalization based on following criteria: probes containing a SNP within 3 bp of the analysed CpG site, bad quality probes based on an iterative greedycut algorithm where a detection p-value of 0.01 was set as a threshold for an unreliable measurement, and probes with missing values in at least one sample. After filtering these cg-probes, beta-values (ratio of methylated probe intensity versus total probe intensity) were within-array normalized using the beta mixture quantile dilation (BMIQ) method [36]. Another filtering step was performed after normalization based on following criteria: probes measuring methylation not at CpG sites and probes on sex chromosomes. The two filtering steps removed a total of 20,338 cg-probes and ended up with a data set of normalized methylation values for 465,239 cg-probes. Beta-values were transformed to M-values ($M = \log_2(\beta/(1-\beta))$) prior to further analyses. Principal component analysis (PCA) was conducted to detect possible batch effects. Associations between the first eight principal components and possible batch effect covariates were measured. The Kruskall-Wallis test was used to find associations with sentrix_ID (BeadChip), while the two-sided Wilcoxon sum rank test was used for associations with the processing date, exposure and *PON1* Q192R genotype. Significant associations between principal component 2 and sentrix_ID (BeadChip) and processing date were suggestive for batch effects and were therefore corrected using the ComBat function in the SVA R package [37] (**Supplementary Figure 1 and 2**). Raw and normalised array data were uploaded to the Gene Expression Omnibus (GEO) database and have accession number: GSE90177.

For each sample the relative cell type contribution was measured using the approach described by Houseman et al. [38]. Reference methylomes of each blood cell type (granulocyte, CD4+ T-cell, CD8+ T-cell, B-cell, monocyte, NK-cell) were obtained from the study of Reinius et al. using the FlowSorted.Blood.450k R package [39]. The analysis was limited to the 100,000 most variable sites. The top 500 cg-probes associated with the cell types were used to estimate the relative cell type composition in each sample. One-way ANOVA was used to determine differences in relative cell type composition between the exposed and unexposed children and between children with the QQ and QR/RR genotype. Associations between relative cell type composition and health outcomes (percentage body fat, delta BMI z-scores from birth to school age, and BMI z-scores), leptin levels and age were analysed using simple linear regression.

Statistical analysis

Differential methylation was analysed both at the single CpG site level and at the region level (**Figure 1**). At the single CpG site level multiple linear regression (Matlab version

2014b, The Mathworks®, Natick, MA, USA) was performed in which methylation was the dependent variable and *PON1* Q192R genotype and prenatal pesticide exposure (yes/no) were the independent variables. Our statistical approach was designed to explain - at the level of methylation - the previously reported gene-environment interaction between the paraoxonase 1 (*PON1*) Q192R genotype and prenatal pesticide exposure leading to an adverse cardio-metabolic risk profile at school age among children carrying the R-allele [19].Thus, our primary interest was to identify methylation marks associated with exposure that were more altered in R-allele carriers than in QQ-homozygotes. Two statistical models were included in our statistical approach. In the first model, effect modification (interaction) of exposure by *PON1* Q192R genotype was allowed by including main effects (exposure and genotype) and cross-product terms (exposure*genotype) in the models. Statistical significant effects of exposure in the *PON1* 192QR/RR group were defined as follows: P-value interaction term ≤ 0.1 and P-value of exposure in the QR/RR group ≤ 0.001. This model allows studying synergistic effects where the combined effect of prenatal exposure and in the QR/RR group is greater than the sum of the effects of each factor alone. In the second model, effect modification of exposure by *PON1* Q192R genotype was not assumed (no cross product term included). Statistical significant effects of exposure were defined as follows: P-value of exposure ≤ 0.001, P-value of *PON1* genotype ≤ 0.1. In this model the combined effect of exposure and being R-allele carrier is equal to the sum of the effect of each factor separately. For both models, the associations were adjusted for child sex. To identify probes that were most aberrant in the exposed QR/RR group, we set an additional filter for both models in which we defined that the prenatally exposed QR/RR group should either be highest or lowest methylated (based on mean methylation level) as compared to the other three groups (exposed QQ, unexposed QR/RR and unexposed QQ). These sites are defined as significantly differentially methylated positions (sig-DMPs) in the remainder of this text. Sig-DMPs were annotated using the HumanMethylation450 v1.2 manifest file. The freely available EpiExplorer tool was used to add further annotation including chromatin state segmentation and histone modifications based on the UCSC hg19 browser [40]. Genomic locations of transcription factor binding sites (TFBS) were directly downloaded from the UCSC h19 genome browser. Enrichment or depletion of sig-DMPs in a particular genomic region was determined using the Fisher's exact test.

DMRs were detected using the limma-based DMRcate R package [41]. We only looked for regions differentially methylated between the exposed QR/RR group and one of the other groups (exposed QQ, unexposed QR/RR and unexposed QQ). In line with identification of sig-DMPs, significant regions (P_{adj}-value < 0.05) were selected in which the exposed R-allele carriers showed either the highest or lowest methylation state

which are called sig-DMRs in the remainder of this text. P-values were corrected for multiple testing using the Benjamini-Hochberg method (P_{adj}).

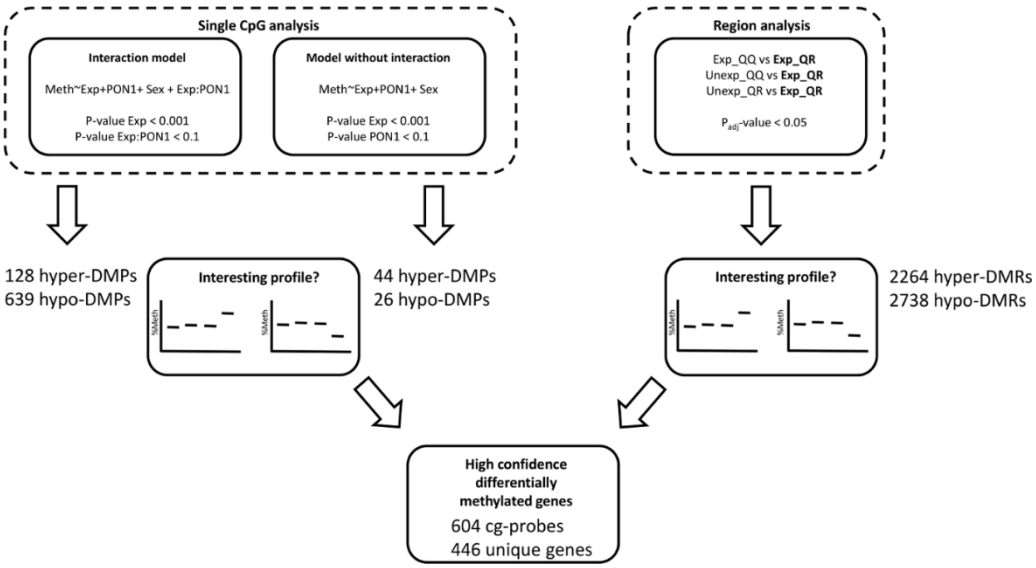


Figure 1: Analysis workflow. Differentially methylated genes were detected using a single CpG and a region-based approach. Only sig-DMPs and sig-DMRs were selected in which the pesticide exposed QR carrier group was either hyper- or hypomethylated in comparison with the other groups (interesting profile). By overlapping the sig-DMPs with the sig-DMRs a high confidence list of differentially methylated genes could be generated.

Pyrosequencing

We used bisulfite pyrosequencing to further verify the methylation differences observed in the methylation array. We selected regions in four genes that are known to be involved in metabolism: *LEP*, *GPR39*, *PPARG* and *OPCML* (**Supplementary Figure 3**). *LEP* DNA methylation has been associated with BMI, birth weight and cholesterol levels [42-44]. Also, maternal conditions have an effect on the methylation status of the *LEP* promoter [45-48]. GPR39 belongs to the ghrelin receptor family and was shown to be associated with obesity [49]. PPARG is a nuclear receptor involved in regulation of lipid and glucose metabolism as well as a target for some obesogenic endocrine disruptors [20, 50-53]. Furthermore, PPARγ is directly involved in the regulation of PON1 gene expression [54-56]. OPCML (Opioid Binding Protein/Cell Adhesion Molecule like) is a member of the IgLON family. A SNP in the OPCML gene was associated with coronary artery calcified plaque in African Americans with type 2 diabetes [57]. A mouse and human GWAS analysis identified an OPCML SNP associated with obesity traits and visceral adipose/subcutaneous adipose ratio, respectively [58, 59]. 1 µg DNA from each sample was bisulfite converted using the EpiTect Fast bisulfite conversion kit

(Qiagen, Hilden, Germany) according to manufacturer's instructions. 15 ng of bisulfite treated DNA was subsequently used in PCR amplification using the PyroMark PCR kit (Qiagen, Hilden, Germany). Reverse primers were biotinylated to get biotin-labelled PCR products. Finally, DNA sequences were pyrosequenced using the PyroMark Q24 Advanced instrument (Qiagen, Hilden, Germany). First, streptavidin-coated Sepharose beads (High Performance, GE Healthcare, Uppsala, Sweden) were used to immobilize the biotin-labelled PCR products. Subsequently, PCR products were captured by the PyroMark vacuum Q24 workstation, washed and denaturated. The single stranded PCR products were mixed and annealed with their corresponding sequencing primer. After the pyrosequencing run was finished, the results were analysed using the PyroMark Q24 Advanced software (Qiagen, Hilden, Germany). Biotinylated-reverse, forward and sequencing primers were designed using the PyroMark assay design 2.0 software (Qiagen, Hilden, Germany) (**Supplementary Table 1**).

Mediation analysis

For a subset of sig-DMPs and sig-DMRs we analyzed 1) whether methylation is a mediator between exposure in *PON1* 192R-allele carriers and leptin levels; and 2) whether methylation is mediator between exposure in *PON1* 192R-allele carriers and body fat accumulation (using delta BMI-score (from birth to school age), and percentage body fat as endpoints). Mediation analysis was restricted to the subset of the methylation data that overlap between the list of sig-DMPs (interaction model) and sig-DMRs. The analysis was performed by the procedure described by Baron and Kenny (1986) [60]. Leptin concentrations were logarithmically (ln) transformed prior to analysis. In mediation analysis considering body fat percentage and leptin, the models were adjusted for sex. As sex was already considered when calculating BMI Z-score, associations considering mediation between pesticide exposure and BMI Z-score were not adjusted for sex.

To demonstrate mediation, four requirements must be met: Model 1) The dependent outcome variable (leptin or a body fat measure) should be significantly associated with pesticide exposure (independent variable); model 2) The DNA methylation mark (mediator) should be significantly associated with pesticide exposure; model 3) The dependent variable should be significantly associated with the DNA methylation mark; and model 4) the DNA methylation mark should be a significant predictor of the outcome variable, while controlling for pesticide exposure. The estimated exposure-related change in the outcome variables in model 4 should be less than in model 1 to demonstrate partial mediation, and drop to zero to demonstrate full mediation. A P-value below 0.05 was used as a cut-off for statistical significance in each of the models.

Functionally relevant mediators, i.e. mediators that have been reported to be involved in development of weight gain/obesity, insulin resistance/diabetes, CVD, and/or fetal growth retardation were subjected to further statistical analysis. R-package "mediation" was used to calculate the significance of the causal mediation effect using a bootstrapping approach [61]. It should be noted that the age of the children varied between 6 and 11 years at the follow-up examination where blood was collected. As child age might affect methylation levels, the exposed and unexposed children selected for this study were age matched within each genotype.

Functional analysis

Ingenuity Pathway Analysis (IPA, Ingenuity Systems®) was used for biological interpretation. The overlap between sig-DMPs and sig-DMRs was determined and used as input for canonical pathway analysis. A Fisher-exact test was used to determine whether the gene lists include more genes associated to a given pathway as compared to random chance (P-value ≤ 0.05).

The DisGeNet platform (http://www.disgenet.org/) was used to screen for gene disease associations [62]. The database (currently) contains 429111 gene disease associations for which the platform provides a reliability score (DisGeNET Score). This score ranges from 0 to 1 and takes into account the number and type of sources (level of curation, organisms), and the number of publications supporting the association (for further details we refer to the DisGeNet website). For this manuscript we extracted the associations with a score above 0.1. By this criterion, 34180 gene disease associations remain in the database. Associated diseases were mapped to the overlapping list of genes between sig-DMPs and sig-DMRs.

Results

Descriptive statistics of the study population

Characteristics, inclusive anthropometric data, for the 48 children (6-11 years of age) are presented in **Table 1**. In accordance with the findings for the whole cohort [19], birth weights were significantly lower and measures of body composition (abdominal circumference, skin fold thickness), increase in BMI Z-score from birth to school age (delta BMI Z-score), diastolic blood pressure, and leptin and insulin concentrations at school age were significantly higher in the exposed *PON1* 192QR/RR group compared with the unexposed QR/RR group. For children with the QQ genotype, none of the variables was significantly affected by prenatal pesticide exposure (p>0.05).

Prenatal pesticide exposure induced methylation changes at CpG sites enriched in promoter regions in *PON1* 192R-allele carriers

Genome-wide DNA methylation in whole blood samples from the children was determined by Illumina 450K methylation arrays and differential methylation patterns related to prenatal pesticide exposure and *PON1* Q192R genotype were analysed. First differential methylation was detected at the single CpG-level using two multiple linear regression models (**Figure 1**). Because relative cell type composition was not associated with pesticide exposure and PON1 Q192R genotype (**Supplementary Figure 4**), differences in cellular composition were not further considered in the workflow of statistical analysis. Allowing effect modification by *PON1* Q192R genotype, 767 sig-DMPs were identified of which 128 were hypermethylated and 639 hypomethylated in prenatally exposed *PON1* 192R allele carriers (**Supplementary Table 2**). When effect modification was not assumed, and the interaction term between exposure and *PON1* genotype was removed from the models, 70 sig-DMPs of which 44 were hypermethylated and 26 hypomethylated in prenatally exposed *PON1* 192R-allele carriers were identified. Hierarchical clustering of the samples using all the sig-DMPs demonstrated a clear cluster of exposed *PON1* 192R-allele carriers (**Figure 2A**). Confidence in detection of differentially methylated genes was increased by further analysis showing that the changes in methylation were not restricted to single CpGs, but were often located in regions or so called DMRs. 5002 sig-DMRs were identified, of which 2264 were hypermethylated and 2738 hypomethylated in the exposed *PON1* 192R carrier group compared to the other groups. Allowing interaction between exposure and *PON1* Q192R genotype to determine sig-DMPs, 547 out of 767 sites (71.3 %) were overlapping with the list of sig-DMRs. When effect modification was not considered, 57 out of 70 sites (81.4 %) were overlapping.

The pyrosequencing methylation percentages confirmed the robustness of Illumina results. They showed significant positive correlations with the Illumina 450K beta-values for all measured CpG probes (**Figure 2B**), except for two probes in the *LEP* gene (cg00840332 and cg26814075) which were borderline significant (p-value: 0.07 and 0.16 respectively). The reason for this less strong correlation between the Illumina and pyrosequencing *LEP* methylation is probably the lower inter-individual methylation variability in this region compared to *GPR39* and *PPARG*.

Table 1. Population characteristics and anthropometric data for 48 pre-pubertal children examined at age 6-11 years stratified by *PON1* Q192R genotype and prenatal pesticide exposure.

	***PON1* 192QQ**		***PON1* QR/RR**	
	unexposed	**exposed**	**unexposed**	**exposed**
N	11	13	12	12
Female sex	5 (45.5)	7 (53.8)	6 (50.0)	6 (50.0)
Maternal smoking in pregnancy	0 (0)	0 (0)	0 (0)	0 (0)
SES[a]	7/4(63.6/36.4)	3/10 (23.1/76.9)*	5/7 (41.7/58.3)	2/10 (16.7/83.3)
Birth weight (g)	3640 (2600; 5412)	3382 (2750; 4573)	3789 (2984; 4345)	3500 (2900; 3914)*
Gestational age (days)	276 (257; 291)	283 (265; 295)	283 (261; 298)	281 (266; 291)
Age (years)	7.6 (6.2; 9.8)	8.4 (6.7; 10,0)	7.8 (6.6; 9.5)	7.7 (7.1; 9.4)
Height (cm)	133.3 (117.3; 145.2)	130.3 (109.7; 139.2)	130.9 (113.7; 149.1)	128.6 (119.3; 142,5)
Weight (kg)	30.9 (18.7; 38.0)	28.3 (18.0; 30.7)	26.3 (19.9; 36.5)	27.4 (19.5; 37.8)
BMI (kg/m²)	16.2 (13.7; 20.5)	15.3 (14.9; 18.3)	15.5 (13.8; 16.9)	15.7 (13.8; 19.7)
BMI Z-scores	0.66 (-1.03; 3.21)	-0.18 (-0.80; 1.49)	-0.04 (-1.31; 0.89)	-0.01 (-0.98; 3.14)
Delta BMI Z-score since birth	-0.45 (-2.15; 2.97)	-0.71 (-2.57; 1.87)	-0.56 (-2.52; 1.03)	0.95 (-2.08; 2.97)*
Abdominal circumference (cm)	60.4 (52.0; 75.8)	58.7 (52.1; 66.8)	58.3 (52.0; 68.1)	60.8 (51.8; 70.6)*
Sum of four skin folds (mm)	38.4 (27.1; 85.4)	33.6 (25.4; 54.5)	34.0 (20.2; 45.2)	44.6 (28.8; 72.0)*
Systolic blood pressure (mmHg)	98.7 (93.7; 110.4)	97.2 (84.3; 105.3)	99.7 (84.7; 106.8)	101.7 (91.0; 108.6)
Diastolic blood pressure (mmHg)	54.7 (46.0; 69.9)	56.2 (46.0; 62.0)	56.3 (49.3; 69.1)	63.0 (57.3; 73.1)**
Leptin (ng/ml)	1.47 (0.70; 9.18)	4.40 (0.60; 15.29)	1.41 (0.67; 5.90)	4.69 (1.79; 12.25)**
Insulin (ng/ml)	0.36 (0.22; 1.15)	0.52 (0.23; 2.55)	0.34 (0.16; 1.62)	1.11 (0.24; 7.10)*
Paraoxonase activity (nmol/min/ml)	27.5 (9.9; 38.0)	30.9 (21.0; 38.9)	58.6 (41.9; 68.7)	59.6 (50.3; 71.5)

Values are presented as median (5-95 percentiles) for continuous variables and as n (%) for categorical variables.

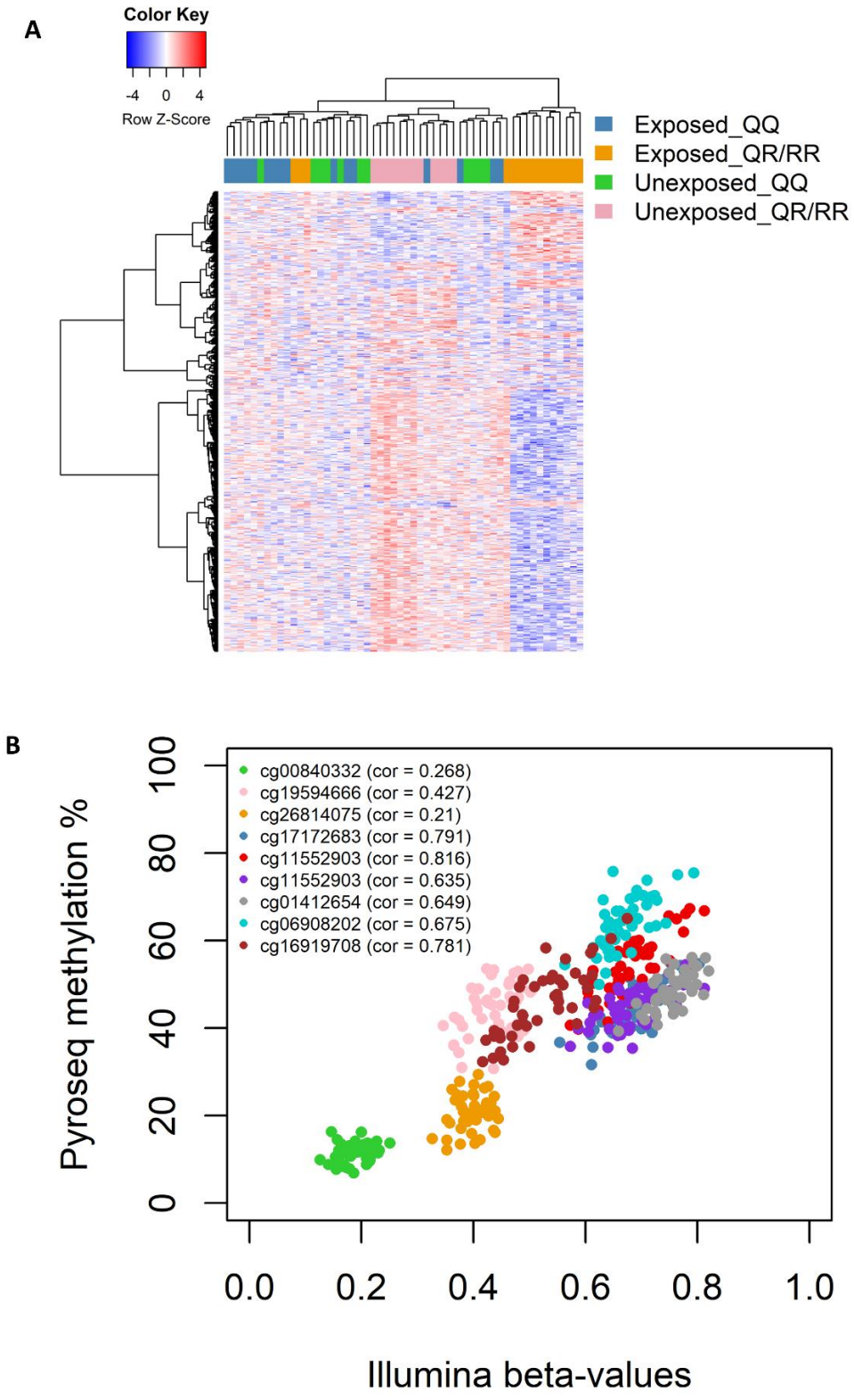
[a]SES: Socioeconomic status (social class 1-3/4-5). Differences between unexposed and exposed children for each *PON1* Q192R genotype were tested using Mann-Whitney U-test for continuous variables and Fishers exact test (dichotomous variables) or Likelihood Ratio (categorical variables with > 2 categories).

* p-value $\leq$ 0.05, ** p-value $\leq$ 0.01

In accordance with the Illumina results, the pyrosequencing *LEP* methylation values were not associated with pesticide exposure and/or *PON1* Q192R genotype. Furthermore, the serum leptin concentrations were not correlated with *LEP* methylation status (data not shown). For *GPR39*, the region analysed with pyrosequencing contained three Illumina cg-probes (cg17172683, cg11552903 and cg18444763), which showed a high correlation (r>0.78) between the Illumina beta-values and pyrosequencing methylation percentages. For most CpGs in the pyrosequencing region we could verify a significant exposure effect and in each CpG site prenatally exposed children with the QR/RR genotype had the lowest mean methylation value (**Supplementary Table 3 and Supplementary Figure 5**). In the *PPARG* promoter a region was selected containing one Illumina cg-probe (cg01412654). Also here the correlation between the 450K Illumina beta-values and the pyrosequencing methylation percentages was strong. However, DNA methylation in this region was not associated with pesticide exposure and/or *PON1* Q192R genotype, and did not correlate with PON1 activity (data not shown). A region in the OPCML gene was found to be higher methylated in prenatal pesticide exposed children carrying the *PON1* 192R-allele. The significant interaction effect between pesticide exposure and *PON1* Q192R genotype could be successfully verified by pyrosequencing. The pyrosequencing methylation values were significant higher methylated in exposed children compared to unexposed children carrying the *PON1* 192R-allele for most of the CpG sites in the region.

Next, we questioned whether the sig-DMPs were enriched or depleted in a specific genomic location (**Figure 2C**). Sig-DMPs for which interaction between exposure and *PON1* Q192R genotype was seen, were enriched in promoter regions (TSS200 and TSS1500, 200 bp and 1500 bp upstream of transcription start sites) and depleted in gene bodies, and 3'UTRs. This was also evident when we overlapped the sig-DMPs with different chromatin states, where we observed enrichment in active and poised promoters, while DMPs were depleted in regions like transcriptional elongation, weak transcribed and heterochromatin regions. Furthermore, DMPs were significantly more located in CpG-islands and less observed in CpG-poor regions. Sig-DMPs found in the models without an interaction term were not found to be enriched or depleted in a particular genomic region.

We also looked for enrichment in TFBS using available chromatin immunoprecipitation (ChIP) ENCODE data from the UCSC genome browser. 39 of the 161 TFBS were significantly enriched for the model with interaction (Bonferroni adjusted P-value < 0.05) while no enrichment was found for the sig-DMPs found in the model without interaction (**Supplementary Table 4**).


A
Color Key
-4 0 4
Row Z-Score
Exposed_QQ
Exposed_QR/RR
Unexposed_QQ
Unexposed_QR/RR
B
Pyroseq methylation %
0 20 40 60 80 100
cg00840332 (cor = 0.268)
cg19594666 (cor = 0.427)
cg26814075 (cor = 0.21)
cg17172683 (cor = 0.791)
cg11552903 (cor = 0.816)
cg11552903 (cor = 0.635)
cg01412654 (cor = 0.649)
cg06908202 (cor = 0.675)
cg16919708 (cor = 0.781)
0.0 0.2 0.4 0.6 0.8 1.0
Illumina beta-values

C

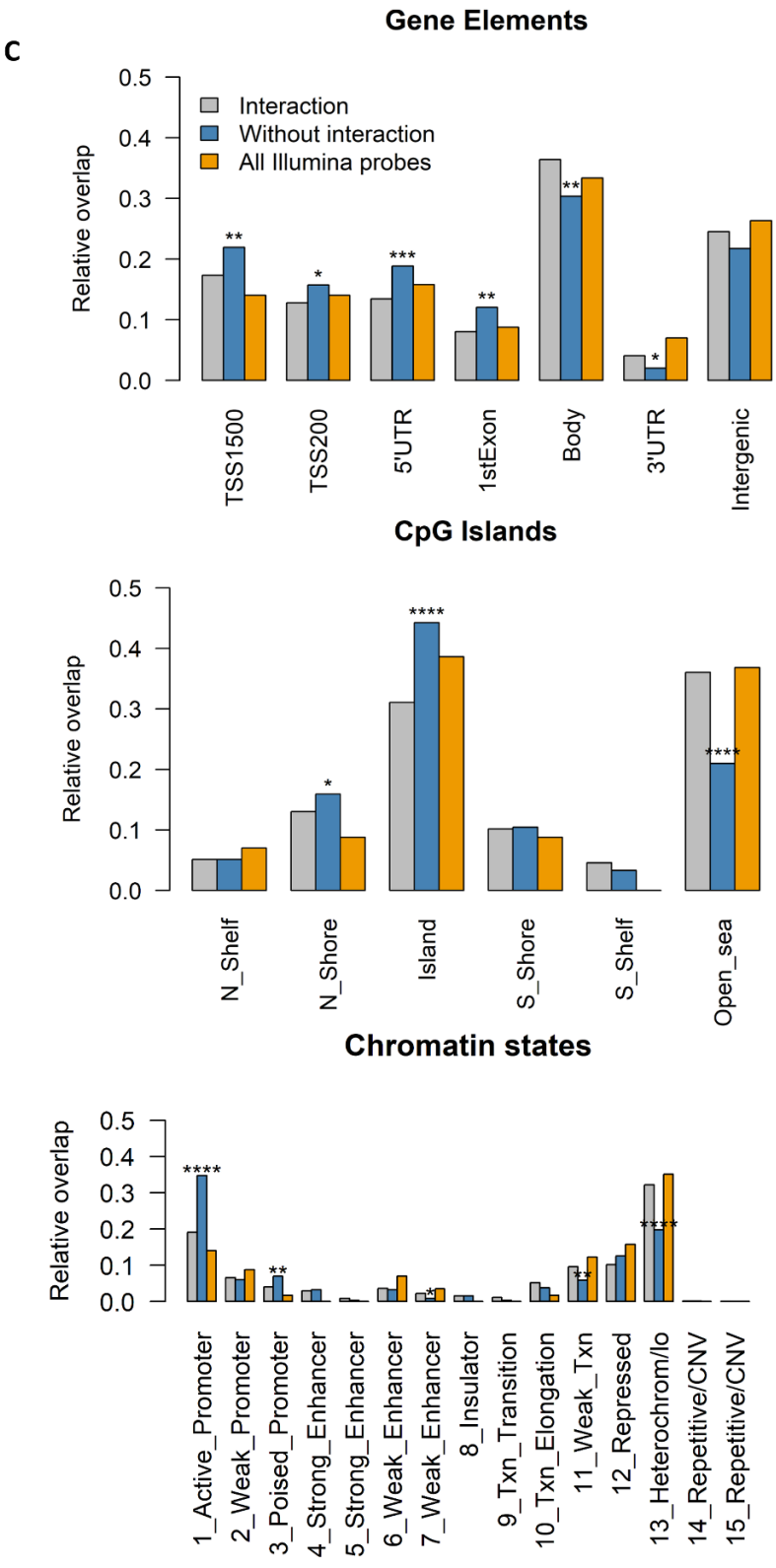

Gene Elements
Interaction
Without interaction
All Illumina probes
Relative overlap
0.0
0.1
0.2
0.3
0.4
0.5
TSS1500
TSS200
5'UTR
1stExon
Body
3'UTR
Intergenic
CpG Islands
N_Shelf
N_Shore
Island
S_Shore
S_Shelf
Open_sea
Chromatin states
1_Active_Promoter
2_Weak_Promoter
3_Poised_Promoter
4_Strong_Enhancer
5_Strong_Enhancer
6_Weak_Enhancer
7_Weak_Enhancer
8_Insulator
9_Txn_Transition
10_Txn_Elongation
11_Weak_Txn
12_Repressed
13_Heterochrom/lo
14_Repetitive/CNV
15_Repetitive/CNV

Figure 2: DNA methylation effect of prenatal pesticide exposure. A) Heatmap of the methylation values from the sig-DMPs showing a clear cluster of prenatal pesticide exposed *PON1*-192 R-carrier samples (orange group). Hierarchical clustering is based on the euclidean distance and average linkage metric. Higher methylation values are colored in yellow, while lower methylation values are colored in blue. **B)** Correlation between the Illumina 450K beta-values and the pyrosequencing methylation percentages. The Pearson correlation coefficient for each CpG probe is indicated between brackets. CpG probes cg00840332, cg19594666 and cg26814075 are located in the LEP gene, cg17172683, cg11552903 and cg18444763 in the GPR39 gene, cg01412654 in the PPARG gene, and cg06908202 and cg16919708 in the OPCML gene. **C)** Genomic location of sig-DMPs. DMPs were mapped to gene elements (top), CGIs (middle) and chromatin state segmentations (bottom). Asterisks indicate significant enrichment or depletion in comparison with all Illumina probes (gray bars) measured by the Fisher's Exact test (P-value < 0.05).

DNA Methylation differences were enriched for genes involved in neuro-endocrine signalling pathways

Overlapping the list of sig-DMPs with the list of sig-DMRs we obtained a robust and high confidence list of differentially methylated genes (N = 446). This list was used as an input for IPA. The top enriched canonical pathways (based on P-value) were dopamine-DARPP32 feedback cAMP signalling, corticotrophin releasing hormone signalling, nNOS signalling in neurons, CDK5 signalling, and neuregulin signalling (**Table 2**). In the context of this manuscript, other significantly enriched pathways such as mTOR signalling (rank 9, -log(P-value) = 1.85) and type II diabetes mellitus signalling (rank 16, -log(P-value) = 1.51) are also highly relevant.

Table 2. Significant enriched Ingenuity canonical pathways.

Rank	Ingenuity canonical pathways	-log(P-value)	Ratio	Hyper-genes	Hypo-genes
1	Dopamine-DARPP32 Feedback in cAMP Signalling	3.98	0.07	*CREB5, PPP2R2B, CACNA1A*	*KCNJ2, NOS1, GRIN2A, GUCY1B3, ADCY2, PRKCH, GNAI3, CACNA1D, PRKCG*
2	Corticotropin Releasing Hormone Signalling	2.74	0.07	*CREB5*	*JUND, NOS1, GUCY1B3, ADCY2, PRKCH, GNAI3, PRKCG*
3	nNOS Signalling in Neurons	2.61	0.11	*CAPN3*	*NOS1, GRIN2A, PRKCH, PRKCG*
4	CDK5 Signalling	2.41	0.07	*PPP2R2B, CACNA1A*	*CDK5R1, NGFR, ITGA2, LAMB1, ADCY2*
5	Neuregulin Signalling	2.06	0.07	*EGFR, ERBB3*	*CDK5R1, ITGA2, PRKCH, PRKCG*
6	PCP pathway	2.06	0.08		*JUND, FZD10, RSPO3, WNT7B, WNT9B*
7	Maturity Onset Diabetes of Young (MODY) Signalling	2.03	0.14	*CACNA1A*	*GAPDH, CACNA1D*
8	Regulation of eIF4 and p70S6K Signalling	2.02	0.05	*PPP2R2B, FAU*	*RPS16, RPS13, RPS10, ITGA2, IRS1, RPS19*
9	mTOR Signalling	1.85	0.05	*PPP2R2B, FAU*	*RPS16, RPS13, RPS10, IRS1, PRKCH, RPS19, PRKCG*
10	Amyotrophic Lateral Sclerosis Signalling	1.84	0.06	*CAPN3, CACNA1A*	*NOS1, GRIN2A, NEFM, CACNA1D*

11	NF-κB Activation by Viruses	1.8	0.07		*ITGAV, CR2, ITGA2, PRKCH, PRKCG*
12	Phosphatidylethanolamine Biosynthesis III	1.7	1		*PTDSS2*
13	Role of CHK Proteins in Cell Cycle Checkpoint Control	1.61	0.07	*PPP2R2B, RFC4*	*E2F3, CHEK1*
14	Synaptic Long Term Depression	1.6	0.05	*IGF1R, PPP2R2B*	*NOS1, GUCY1B3, PRKCH, GNAI3, PRKCG*
15	ErbB Signalling	1.53	0.06	*EGFR, ERBB3*	*NCK2, PRKCH, PRKCG*
16	Type II Diabetes Mellitus Signalling	1.51	0.05	*PKM*	*NGFR, ADIPOR2, IRS1, PRKCH, PRKCG*
17	G Beta Gamma Signalling	1.49	0.06	*EGFR*	*ADCY2, PRKCH, GNAI3, PRKCG*
18	p70S6K Signalling	1.48	0.05	*EGFR, PPP2R2B*	*IRS1, PRKCH, GNAI3, PRKCG*
19	Role of Osteoblasts, Osteoclasts and Chondrocytes in Rheumatoid Arthritis	1.47	0.04		*FZD10, NGFR, SMAD5, WNT7B, ITGA2, IL1RAP, WNT9B, TCF7L2, NFATC1*
20	Molecular Mechanisms of Cancer	1.46	0.04		*RASGRF1, ITGA2, WNT7B, IRS1, E2F3, GNAI3, FZD10, SMAD5, ADCY2, WNT9B, PRKCH, CHEK1, PRKCG*
21	nNOS Signalling in Skeletal Muscle Cells	1.45	0.13	*CAPN3*	*NOS1*
22	Factors Promoting Cardiogenesis in Vertebrates	1.42	0.05		*FZD10, SMAD5, PRKCH, TCF7L2, PRKCG*
23	RAR Activation	1.41	0.04		*REL, ERCC2, SMAD5, NR2F1, ADCY2, PRKCH, RARB, PRKCG*
24	Choline Degradation I	1.4	0.5	*CHDH*	
25	Sulfate Activation for Sulfonation	1.4	0.5	*PAPSS2*	
26	Mismatch Repair in Eukaryotes	1.4	0.13	*RFC4*	*MLH1*
27	Glioma Signalling	1.37	0.05	*IGF1R, EGFR*	*PRKCH, E2F3, PRKCG*
28	Netrin Signalling	1.36	0.08		*UNC5C, NCK2, NFATC1*
29	Cellular Effects of Sildenafil (Viagra)	1.33	0.05	*CACNG6, CACNA1A*	*KCNN1, GUCY1B3, ADCY2, CACNA1D*
30	GNRH Signalling	1.33	0.05	*EGFR, CREB5*	*ADCY2, PRKCH, GNAI3, PRKCG*
31	Protein Kinase A Signalling	1.31	0.03	*HIST1H1A, CREB5*	*PTPN9, TIMM50, NFATC1, GNAI3, AKAP12, NGFR, PTP4A1, ADCY2, PRKCH, TCF7L2, PRKCG*
32	Ovarian Cancer Signalling	1.31	0.05	*EGFR*	*FZD10, WNT7B, MLH1, WNT9B, TCF7L2*
33	Colorectal Cancer Metastasis Signalling	1.3	0.04	*EGFR*	*ADRBK1, APPL1, FZD10, WNT7B, MLH1, ADCY2, WNT9B, TCF7L2*
34	Agrin Interactions at Neuromuscular Junction	1.3	0.06	*EGFR, ERBB3*	*ITGA2, LAMB1*
35	Growth Hormone Signalling	1.3	0.06	*IGF1R*	*IRS1, PRKCH, PRKCG*

DNA methylation (partially) mediates associations between pesticide exposure and higher leptin concentrations, body fat content, and delta BMI z-scores

The list of genes that overlaps between sig-DMPs (as identified by the interaction model) and sig-DMRs was also used as input for mediation analysis. We identified respectively 20, 31, and 45 candidate methylation marks that (partially) mediate the effect between pesticide exposure and serum leptin concentrations; delta BMI Z-score; and body fat content (**Supplementary Table 5**). Based on applied cut-off criteria, we were not able to identify methylation marks that mediate the effect on BMI Z-score. Currently known gene disease associations allowed to extract mediators that were reported to be involved in development of weight gain/obesity, insulin resistance/diabetes, CVD, and/or fetal growth retardation. This subset of mediators is given in **Table 3**. Based on Baron and Kenny's steps to analyze mediation, the association between pesticide exposure and delta BMI Z-score was partially mediated by hypomethylation of *UQCRC2*, *MTNR1B* and *GRIN2A*, and by hypermethylation of *FABP4* and *LRP8*. Methylation of *UQCRC2* and *LRP8* was also a partial mediator in the association between pesticide exposure and body fat percentage. *LRP8* was also found to mediate the association between pesticide exposure and serum leptin concentration. The p-value for significance of the causal mediation effect is included in table 3 and was below 0.1 for all mediators except for *UQCRC2* and *GRIN2A*.

Table 3. Methylation marks that partially mediate the association between pesticide exposure and leptin and body fat accumulation in *PON1*-192 R-allele carriers.

Outcome	IlmnID	nearest Gene Symbol	Gene Name	Direction of methylation in Exposed R carriers	Diseases	Significance of causal mediation effect (p-value)
Leptin	cg03366858	*LRP8*	low density lipoprotein receptor-related protein 8, apolipoprotein e receptor	HYPER	**Myocardial Infarction (0.22)** | Nerve Degeneration (0.21) | **MYOCARDIAL INFARCTION, SUSCEPTIBILITY TO, 1 (finding) (0.2)**	0.02
Leptin	cg18202502	*LRP8*	low density lipoprotein receptor-related protein 8, apolipoprotein e receptor	HYPER	**Myocardial Infarction (0.22)** | Nerve Degeneration (0.21) | **MYOCARDIAL INFARCTION, SUSCEPTIBILITY TO, 1 (finding) (0.2)**	0.024
Delta BMI Z-score	cg00810945	*UQCRC2*	ubiquinol-cytochrome c reductase	HYPO	MITOCHONDRIAL COMPLEX III DEFICIENCY, NUCLEAR	0.138

			core protein II		TYPE 5 (0.41) | **Obesity (0.21)**	
Delta BMI Z-score	cg06337557	*MTNR1B*	melatonin receptor 1B	HYPO	**Diabetes Mellitus, Type 2 (0.26)** | Polycystic Ovary Syndrome (0.21) | Child Development Disorders, Pervasive (0.21) | Acute pancreatitis (0.1)	0.032
Delta BMI Z-score	cg14152613	*FABP4*	fatty acid binding protein 4, adipocyte	HYPER	Carcinoma (0.21) | Mammary Neoplasms, Experimental (0.21) | Mammary Neoplasms, Animal (0.21) | **Insulin Resistance (0.1)** | Erectile Dysfunction (0.1) | **Diabetes Mellitus, Experimental (0.1)**	0.068
Delta BMI Z-score	cg15134033	*GRIN2A*	glutamate receptor, ionotropic, N-methyl D-aspartate 2A	HYPO	Epilepsy (0.21) | Colorectal Neoplasms (0.21) | Epilepsy, Rolandic (0.21) | Melanoma (0.21) | Landau-Kleffner Syndrome (0.21) | Autistic Disorder (0.21) | Morphine Dependence (0.21) | Language Development Disorders (0.21) | EPILEPSY, FOCAL, WITH SPEECH DISORDER AND WITH OR WITHOUT MENTAL RETARDATION (0.21) | Speech Disorders (0.21) | Substance Withdrawal Syndrome (0.21) | Rolandic Epilepsy, Mental Retardation, And Speech Dyspraxia, Autosomal Dominant (0.2) | Reperfusion Injury (0.1) | Hypoxia-Ischemia, Brain (0.1) | Sepsis (0.1) | **Fetal Growth Retardation (0.1)** | Central Nervous System Viral Diseases (0.1) | Placental Insufficiency (0.1)	0.144

Delta BMI Z-score	cg18202502	*LRP8*	low density lipoprotein receptor-related protein 8, apolipoprotein e receptor	HYPER	**Myocardial Infarction (0.22)** | Nerve Degeneration (0.21) | **MYOCARDIAL INFARCTION, SUSCEPTIBILITY TO, 1 (finding) (0.2)**	0.026
Bodyfat	cg00810945	*UQCRC2*	ubiquinol-cytochrome c reductase core protein II	HYPO	MITOCHONDRIAL COMPLEX III DEFICIENCY, NUCLEAR TYPE 5 (0.41) | **Obesity (0.21)**	0.174
Bodyfat	cg03366858	*LRP8*	low density lipoprotein receptor-related protein 8, apolipoprotein e receptor	HYPER	**Myocardial Infarction (0.22)** | Nerve Degeneration (0.21) | **MYOCARDIAL INFARCTION, SUSCEPTIBILITY TO, 1 (finding) (0.2)**	< 0.001
Bodyfat	cg18202502	*LRP8*	low density lipoprotein receptor-related protein 8, apolipoprotein e receptor	HYPER	**Myocardial Infarction (0.22)** | Nerve Degeneration (0.21) | **MYOCARDIAL INFARCTION, SUSCEPTIBILITY TO, 1 (finding) (0.2)**	< 0.001

Only the subset of genes for which associations with metabolic disease have been reported is listed. DisGeNET Score – indicating reliability of the gene disease associations - is included between brackets.

DNA methylation at the *PON1* promoter is affected by the *PON1* -108CT SNP (rs705379) and negatively correlated with paraoxonase 1 activity

Beside the genome-wide DNA methylation effects of the *PON1* Q192R genotype, we also observed a wide variation in DNA methylation in the *PON1* promoter itself for nine Illumina cg-probes. Prenatal pesticide exposure and/or *PON1* Q192R genotype did not affect *PON1* promoter methylation status. However, another polymorphism (rs705379, *PON1* -108CT) in the promoter region of PON1 could explain a large extent of this variation (**Figure 3**). Individuals homozygous for the T-allele showed higher methylation values compared with the homozygous C-allele carriers. As expected, heterozygous individuals had an intermediate methylation value. Furthermore, the paraoxonase 1 activity was significantly associated with DNA methylation in the *PON1* promoter region, with higher methylation values resulting in lower paraoxonase 1 activity (**Figure 4**). *PON1* Q192R genotype had the strongest effect on PON1 activity, while variation in PON1 promoter methylation led to a smaller but significant effect on PON1 activity.

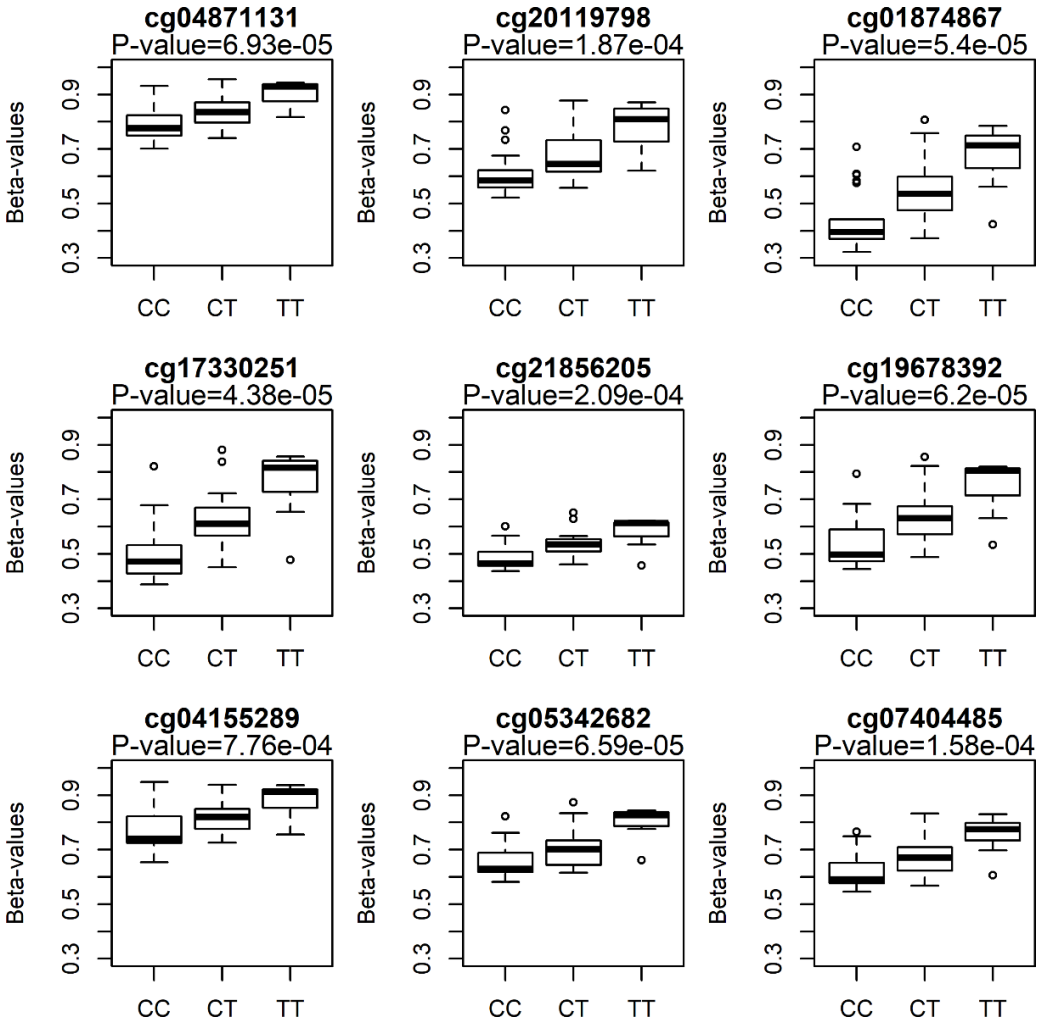


Figure 3: Association between *PON1* methylation and *PON1* C-108T SNP. Individuals homozygous for the T allele showed higher methylation values (beta-values) as compared with C allele carriers. P-values shown are those from the one-way ANOVA analysis.

Discussion

We found that prenatal pesticide exposure was associated with a differential DNA methylation profile in children carrying the *PON1* 192R-allele compared to children with the *PON1* 192QQ genotype and unexposed children. 767 sig-DMPs were identified of which 128 were hypermethylated and 639 hypomethylated in prenatally exposed *PON1* 192R-allele carriers. The profiles of *PON1* 192R-allele carriers clustered together. As far as we know, our study is the first one to demonstrate a link between epigenetics and genetic susceptibility towards pesticide exposure in fetal life. Our study supports a

linkage of a differential methylation pattern and higher body fat content and serum leptin concentrations in school age children dependent on both *PON1* Q192R genotype and prenatal pesticide exposure.

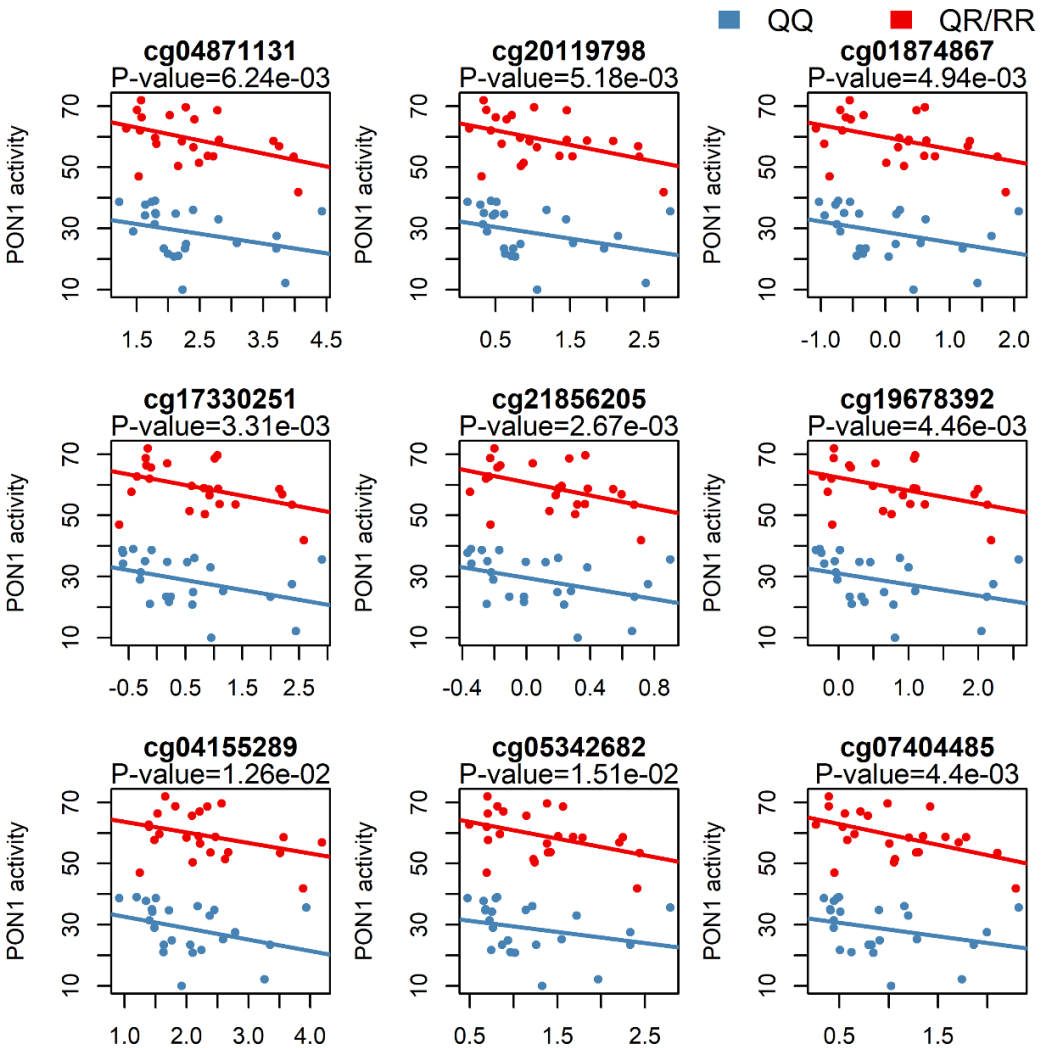


Figure 4: Association between PON1 activity and *PON1* methylation. The P-values of the main effect for methylation are displayed using the linear model PON1 activity ~ M-value + *PON1*-192 genotype + sex. Red colored samples are *PON1* 192 R-allele carriers and samples in blue are children with the *PON1* 192QQ genotype.

The majority of the detected sig-DMPs were hypomethylated in exposed children with the *PON1* 192QR/RR genotype. Interestingly, these DMPs were mainly located in gene promoters, CpG-islands and TFBSs, suggesting a possible direct link with gene expression. To increase the confidence of our findings we also screened for DMRs. Most

of the single CpG-sites were part of a DMR suggesting that these were independent of technical variation and could be considered as reliable.

Technical reliability of the outcomes from the 450K Illumina methylation array was successfully confirmed by bisulfite pyrosequencing of corresponding CpG probe regions of four selected genes, i.e. *LEP*, *PPARG, GPR39* and *OPCML* for which corresponding probes were available.

LEP was chosen because we previously found leptin to be a potential mediator of the association between prenatal pesticide exposure and body fat accumulation in children with the *PON1* 192R-allele [20]. In addition, multiple studies demonstrated associations between *LEP* DNA methylation and BMI, birth weight and cholesterol concentrations [42-44]. LEP was also found to be differentially methylated in the offspring of mothers suffering from the Dutch winter famine [45]. However, our pyrosequencing results did not demonstrate a correlation between leptin DNA methylation and leptin serum concentrations, and prenatal pesticide exposure was not associated with changes in leptin DNA methylation. This suggests that the higher leptin concentration observed in exposed children with the R-allele is not due to a direct effect on DNA methylation of the leptin gene itself. Another gene whose methylation was confirmed by pyrosequencing was *PPARG*, a nuclear receptor controlling the expression of genes involved in lipid storage and glucose metabolism and target for obesogenic compounds [50-53]. Furthermore, PPARγ is involved in the regulation of *PON1* expression [54-56]. However, we did not find a correlation between *PPARG* DNA methylation and PON1 activity (data not shown). In our dataset, prenatal pesticide exposure did not seem to change *PPARG* methylation levels irrespective of *PON1* Q192R genotype.

Reduced *GPR39* DNA methylation observed in prenatally pesticide exposed R-allele carriers was confirmed with pyrosequencing. GPR39 is receptor for obestatin (belonging to the ghrelin receptor family), involved in regulation of appetite and glucose homeostasis [63, 64] and associated with obesity [49]. Furthermore, GPR39 knock-out mice showed an increased fat accumulation due to changes in lipolysis and energy expenditure [49]. So, mis-regulation of this gene due to methylation changes might lead to an obese phenotype. To our knowledge, no other study has yet reported methylation differences in this region associated with obesity or metabolic disorders, or showed links with pesticide exposure.

The higher methylation values of the *OPCML* DMR in exposed children carrying the *PON1* 192R-allele could be confirmed by pyrosequencing. *OPCML* encodes for a protein belonging the IgLON family. OPCML was shown to be a tumor suppressor and inactivated by DNA methylation in a variety of cancer types [65-68]. There is also a link with metabolic diseases, as SNPs in this gene were found to be associated with obesity

traits, coronary artery calcified plaque and visceral adipose/subcutaneous adipose ratio [57-59].

Further analysis revealed that the differences in DNA methylation were most pronounced in genes involved in neuro-endocrine signalling pathways, including "dopamine-DARPP32 feedback in cAMP signalling", "corticotropin releasing hormone signalling", "nNOS signalling in neurons", and "CDK5 signalling". These pathways are important in the control of food intake and energy balance. Dopamine signalling, for example, is one of the key players in the reward pathway, also controlling food intake and preferences. Reduced dopamine signalling is assumed to induce overeating [69, 70]. In mice, a high-fat diet during pregnancy resulted in altered gene expression and DNA methylation of the dopamine transporter gene in the offspring, leading to an increased preference for sucrose and fat [71]. Another study found similar results in prenatally stressed rats given a high fat-sucrose diet [72]. These studies suggest that prenatal and early life conditions may influence food intake and food preferences later in life through modulation of the dopamine pathway [73-77]. Organophosphate insecticides have been shown to modulate dopamine signalling [78]. Furthermore, low-dose exposure of neonatal rats caused metabolic dysfunction resembling prediabetes and in adulthood exposed animals gained excess weight when fed a high fat diet compared to unexposed rats on the same diet [79].

Corticotropin-releasing hormone (CRH) is a neuropeptide secreted in response to stress. However, a role for CRH in regulating energy balance and food intake has also been described [80-82] including a relation to the action of leptin [83].

Also NOS1 neurons are involved in energy balance and food intake [84-86]. Knock-out of *NOS1* in leptin receptor- and NOS1-expressing hypothalamic neurons results in hyperphagic obesity, decreased energy expenditure and hyperglycemia in mice [85]. Interestingly, organophosphates have been shown to alter NOS1-expressing neurons during development in mice [87, 88].

Neureguline 1 treatment in rodents has been shown to increase serum leptin concentrations, prevent weight gain and lower food intake. Hence, affecting this pathway may also change food intake and energy metabolism [89, 90].

A limitation of this study is that the methylation profile is measured at the same time as health outcomes and causality as such cannot be proven. Some of the genes that relate to the sig-DMPs are involved in neuro-endocrine pathways that regulate appetite and energy balance but this study cannot rule out if these sig-DMPs are a consequence of alterations of food habits and physical activity among the exposed children with the *PON1* 192R-allele or an underlying mechanism. However, the mediation analysis suggested that some of the differentially methylated marks are on the mechanistic pathway between prenatal pesticide exposure and the measured outcomes. This result

suggests that, at least in some, CpG-sites a change in methylation might contribute to metabolic disturbances later in life. Furthermore, the association was not significant between pesticide exposure and BMI Z-score as such, but between pesticide exposure and delta BMI Z-score which integrates fat accumulation from birth and onwards to school age.

Interestingly, some of the mediator marks could be linked to specific genes that were reported earlier to play a role in the development of weight gain/obesity, insulin resistance/diabetes, CVD, and/or fetal growth retardation: *UQCRC2*, *MTNR1B*, *GRIN2A*, *FABP4* and *LRP8*. *FABP4* encodes for a member of the fatty acid binding protein family regulating lipid trafficking, signalling and metabolism. Different studies have demonstrated the role of this protein in obesity, type 2 diabetes and atherosclerosis development [91-93]. In ApoE deficient mice with hyperhomocysteine *FABP4* DNA methylation is reduced in the aorta compared to wild type mice, leading to a higher gene expression [94, 95]. *UQCRC2* encodes a protein which is part of the ubiquinol-cytochrome c reductase complex in the mitochondria. UQCRC2 was shown to be downregulated in individuals who were susceptible to weight gain and obesity development [96]. The melatonin receptor 1B (MTNR1B) has a main function in regulating circadian rhythm. Interestingly, several polymorphisms in the *MTNR1B* gene are associated with type 2 diabetes, fasting glucose concentration and insulin secretion [97-99]. *GRIN2A* encodes for a NMDA glutamate receptor subunit. Polymorphisms in the *GRIN2A* gene are associated with epilepsy and different neurological and mental disorders [100-104]. A decreased gene expression of GRIN2A in rats after intrauterine growth retardation suggests a possible role for this gene in fetal growth and development [105]. *LRP8* encodes for a member of the LDL receptor family. Common polymorphisms in the *LRP8* gene are associated with coronary artery disease, myocardial infarction and high birth weight [106-110]. Thus, the mediation analysis suggests a mechanistic role of epigenetics in the development of an adverse metabolic risk profile among the prenatally exposed children with the *PON1* R-allele as previously reported for these children [19] and confirmed in the selected subset of children.

A few studies have investigated associations between PON1 genotype and metabolic disturbances in children. A recent study showed a higher risk of insulin resistance (HOMA-IR) in Mexican children with the RR-genotype as compared to children with the QQ or QR genotypes although BMI did not differ between the groups [111]. Among Mexican-American children from an agricultural community in California a trend of increased BMI Z-scores with increased number of PON1 192Q alleles was seen [112]. However, potential interactions between PON1 genotype and prenatal exposure to pesticides, or other environmental contaminants, were not investigated in these studies. In our cohort, unexposed QQ-homozygote children also tended to have higher

body fat content than unexposed R-carriers but prenatally pesticide exposed children with the R-allele accumulated more fat during childhood and had a more unhealthy metabolic risk profile at school age than unexposed children and exposed children with the QQ genotype [19].

We also demonstrated that methylation in the *PON1* promoter itself is affected by a SNP (*PON1* -108CT, rs705379). In addition, *PON1* methylation values were negatively associated with paraoxonase 1 activity. These results are in agreement with the outcome of a recent study from Huen and colleagues [113]. They found methylation in the same nine CpG sites to be associated with the *PON1* -108CT polymorphism and also reported an inverse association with arylesterase (AREase) activity as a measure of *PON1* expression, both in newborns and 9-years old children. Furthermore, they demonstrated that *PON1* methylation mediates the relationship between *PON1* expression and the promoter -108 genotype. However, the effect of prenatal pesticide exposure on the health outcomes shown in table 1 was not modulated by PON1 -108CT genotype (data not shown).

Our findings indicate that the higher vulnerability among children with the R-allele towards prenatal pesticide exposure might be mediated by genotype-specific epigenetic alterations. However, a limitation of this study is that we cannot identify individual pesticides related to these findings, since the study design did not allow bio-monitoring of pesticide exposure in the mothers, and the exposure classification of the mothers encompassed more than 100 pesticides used in different mixtures [11].

However, the existence of mixed exposure is a real-world situation, and the longitudinal design, the blinded exposure classification, and the blinded clinical examinations and genotyping minimized the possible impact of exposure misclassification and bias.

Since PON1 is known to detoxify some organophosphate insecticides (e.g., chlorpyrifos) and these substances were frequently applied in the mothers working areas, organophosphate insecticides could be assumed to be responsible for the observed effects. However, the mechanism is unclear and does not seem to be related to the hydrolysis efficiency, since R-carriers have higher paraoxonase activity than QQ homozygotes. Besides, at relatively low exposure levels, as in this study, the capacity to detoxify organophosphates is considered to be independent of the *PON1* Q192R genotype [114] and furthermore serum PON1 activity was reported to be low in newborns and may be even lower before birth, as indicated by lower activity in premature compared to term babies [115, 116]. Thus, differences in fetal detoxification of pesticides related to *PON1* genotype might not be a likely explanation of the exposure-related difference in methylation pattern between children with the QR/RR and QQ genotype.

Another limitation of the study is that DNA methylation analyses were performed in white blood cells as surrogates for the target tissues. We do not know whether the differences in DNA methylation patterns found in blood mirror a similar change in adipose tissue for example. A recent study from Huang et al. demonstrated several potential limitations in using methylation profiles in blood to mirror the corresponding profile in target tissues by comparing paired blood and adipose tissue methylation profiles [117]. Furthermore, the composition of blood cell types may be variable and might affect the DNA methylation analyses. In our dataset prenatal pesticide exposure and/or *PON1* Q192R genotype did not affect the relative blood cell counts determined by the reference-based method of Houseman. Cell counts were not included in the models due to the small sample size of the study. Since we found that some of the health effects (mainly leptin) were associated with cell type count (**Supplementary Table 6**), we cannot exclude that the results of the mediation analysis were biased by differences in cell type composition. Based on the data of Reinius et al. [39], methylation of only two CpG probes (cg18202502 and cg15134033) in table 3 were slightly associated with cell types (data not shown). Methylation in the other CpG probes in table 3 were not significantly different between the blood cell types.

Finally, the small number of subjects included in this exploratory study is a clear limitation because of the limited statistical power. Despite these limitations, our findings suggest that DNA methylation might be a link between prenatal pesticide exposure and a cardio-metabolic risk profile in children carrying the PON1 192R-allele. The findings deserve further investigation in a larger study with quantitative data on pesticide exposure. Whether this DNA methylation pattern is unique to pesticide exposure or is shared by other adverse prenatal environmental factors also needs further investigation.

In summary, our data indicate that DNA methylation may be an underlying mechanism explaining an adverse cardio-metabolic risk profile in prenatally pesticide exposed children carrying the *PON1* 192R-allele.

References

1. Bjorling-Poulsen, M., H.R. Andersen, and P. Grandjean, *Potential developmental neurotoxicity of pesticides used in Europe.* Environ Health, 2008. **7**: p. 50.
2. Andersen, H.R., et al., *Effects of currently used pesticides in assays for estrogenicity, androgenicity, and aromatase activity in vitro.* Toxicol Appl Pharmacol, 2002. **179**(1): p. 1-12.
3. Orton, F., et al., *Widely Used Pesticides with Previously Unknown Endocrine Activity Revealed as in Vitro Anti-Androgens.* Environ.Health Perspect., 2011. **119**(6): p. 794-800.
4. Grandjean, P. and P.J. Landrigan, *Neurobehavioural effects of developmental toxicity.* Lancet Neurol, 2014. **13**(3): p. 330-8.
5. London, L., et al., *Neurobehavioral and neurodevelopmental effects of pesticide exposures.* Neurotoxicology, 2012. **33**(4): p. 887-96.
6. Li, A.A., et al., *Building a scientific framework for studying hormonal effects on behavior and on the development of the sexually dimorphic nervous system.* Neurotoxicology, 2008. **29**(3): p. 504-519.
7. Gore, A.C., *Neuroendocrine targets of endocrine disruptors.* Hormones (Athens), 2010. **9**(1): p. 16-27.
8. Jacobsen, P.R., et al., *Combined exposure to endocrine disrupting pesticides impairs parturition, causes pup mortality and affects sexual differentiation in rats.* Int J Androl, 2010. **33**(2): p. 434-42.
9. Wohlfahrt-Veje, C., et al., *Lower birth weight and increased body fat at school age in children prenatally exposed to modern pesticides: a prospective study.* Environ Health, 2011. **10**(1): p. 79.
10. Wohlfahrt-Veje, C., et al., *Smaller genitals at school age in boys whose mothers were exposed to non-persistent pesticides in early pregnancy.* Int J Androl, 2012. **35**(3): p. 265-72.
11. Andersen, H.R., et al., *Impaired reproductive development in sons of women occupationally exposed to pesticides during pregnancy.* Environ Health Perspect, 2008. **116**(4): p. 566-72.
12. Wohlfahrt-Veje, C., et al., *Early breast development in girls after prenatal exposure to non-persistent pesticides.* Int J Androl, 2012. **35**(3): p. 273-82.
13. Andersen, H.R., et al., *Occupational pesticide exposure in early pregnancy associated with sex-specific neurobehavioral deficits in the children at school age.* Neurotoxicol Teratol, 2015. **47**: p. 1-9.
14. Costa, L.G., T.B. Cole, and C.E. Furlong, *Polymorphisms of paraoxonase (PON1) and their significance in clinical toxicology of organophosphates.* J.Toxicol.Clin.Toxicol., 2003. **41**(1): p. 37-45.
15. Mackness, B., et al., *Paraoxonase and susceptibility to organophosphorus poisoning in farmers dipping sheep.* Pharmacogenetics, 2003. **13**(2): p. 81-88.
16. Aviram, M., et al., *Paraoxonase inhibits high-density lipoprotein oxidation and preserves its functions. A possible peroxidative role for paraoxonase.* J.Clin.Invest, 1998. **101**(8): p. 1581-1590.
17. Durrington, P.N., B. Mackness, and M.I. Mackness, *Paraoxonase and atherosclerosis.* Arterioscler.Thromb.Vasc.Biol., 2001. **21**(4): p. 473-480.
18. Seo, D. and P. Goldschmidt-Clermont, *The paraoxonase gene family and atherosclerosis.* Curr.Atheroscler.Rep., 2009. **11**(3): p. 182-187.
19. Andersen, H.R., et al., *Paraoxonase 1 polymorphism and prenatal pesticide exposure associated with adverse cardiovascular risk profiles at school age.* PLoS One, 2012. **7**(5): p. e36830.
20. Jorgensen, A., et al., *Interaction between paraoxonase 1 polymorphism and prenatal pesticide exposure on metabolic markers in children using a multiplex approach.* Reprod Toxicol, 2015. **51**: p. 22-30.
21. Patel, S.B., et al., *Leptin: linking obesity, the metabolic syndrome, and cardiovascular disease.* Curr Hypertens Rep, 2008. **10**(2): p. 131-7.
22. Meas, T., et al., *Plasminogen activator inhibitor type-1 is an independent marker of metabolic disorders in young adults born small for gestational age.* J Thromb Haemost, 2010. **8**(12): p. 2608-13.

23. Huang, K.C., et al., *Plasma leptin is associated with insulin resistance independent of age, body mass index, fat mass, lipids, and pubertal development in nondiabetic adolescents.* Int J Obes Relat Metab Disord, 2004. **28**(4): p. 470-5.
24. Ho, S.M., et al., *Environmental epigenetics and its implication on disease risk and health outcomes.* ILAR J, 2012. **53**(3-4): p. 289-305.
25. Faulk, C. and D.C. Dolinoy, *Timing is everything: the when and how of environmentally induced changes in the epigenome of animals.* Epigenetics, 2011. **6**(7): p. 791-7.
26. Saffery, R. and B. Novakovic, *Epigenetics as the mediator of fetal programming of adult onset disease: what is the evidence?* Acta Obstet Gynecol Scand, 2014. **93**(11): p. 1090-8.
27. Reynolds, R.M., G.H. Jacobsen, and A.J. Drake, *What is the evidence in humans that DNA methylation changes link events in utero and later life disease?* Clin Endocrinol (Oxf), 2013. **78**(6): p. 814-22.
28. Perera, F. and J. Herbstman, *Prenatal environmental exposures, epigenetics, and disease.* Reprod Toxicol, 2011. **31**(3): p. 363-73.
29. Chmurzynska, A., *Fetal programming: link between early nutrition, DNA methylation, and complex diseases.* Nutr Rev, 2010. **68**(2): p. 87-98.
30. Knopik, V.S., et al., *The epigenetics of maternal cigarette smoking during pregnancy and effects on child development.* Dev Psychopathol, 2012. **24**(4): p. 1377-90.
31. Casati, L., et al., *Endocrine disrupters: the new players able to affect the epigenome.* Front Cell Dev Biol, 2015. **3**: p. 37.
32. Skinner, M.K., *Environmental epigenomics and disease susceptibility.* EMBO Rep, 2011. **12**(7): p. 620-2.
33. Relton, C.L., et al., *DNA methylation patterns in cord blood DNA and body size in childhood.* PLoS One, 2012. **7**(3): p. e31821.
34. Godfrey, K.M., et al., *Epigenetic gene promoter methylation at birth is associated with child's later adiposity.* Diabetes, 2011. **60**(5): p. 1528-34.
35. Assenov, Y., et al., *Comprehensive analysis of DNA methylation data with RnBeads.* Nat Methods, 2014. **11**(11): p. 1138-1140.
36. Teschendorff, A.E., et al., *A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data.* Bioinformatics, 2013. **29**(2): p. 189-96.
37. Leek, J.T., et al., *The sva package for removing batch effects and other unwanted variation in high-throughput experiments.* Bioinformatics, 2012. **28**(6): p. 882-3.
38. Houseman, E.A., et al., *DNA methylation arrays as surrogate measures of cell mixture distribution.* BMC Bioinformatics, 2012. **13**: p. 86.
39. Reinius, L.E., et al., *Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility.* PLoS One, 2012. **7**(7): p. e41361.
40. Halachev, K., et al., *EpiExplorer: live exploration and global analysis of large epigenomic datasets.* Genome Biol, 2012. **13**(10): p. R96.
41. Peters, T.J., et al., *De novo identification of differentially methylated regions in the human genome.* Epigenetics Chromatin, 2015. **8**: p. 6.
42. Obermann-Borst, S.A., et al., *Duration of breastfeeding and gender are associated with methylation of the LEPTIN gene in very young children.* Pediatr Res, 2013. **74**(3): p. 344-9.
43. Garcia-Cardona, M.C., et al., *DNA methylation of leptin and adiponectin promoters in children is reduced by the combined presence of obesity and insulin resistance.* Int J Obes (Lond), 2014. **38**(11): p. 1457-65.
44. Houde, A.A., et al., *Leptin and adiponectin DNA methylation levels in adipose tissues and blood cells are associated with BMI, waist girth and LDL-cholesterol levels in severely obese men and women.* BMC Med Genet, 2015. **16**: p. 29.
45. Tobi, E.W., et al., *DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific.* Hum Mol Genet, 2009. **18**(21): p. 4046-53.
46. Jousse, C., et al., *Perinatal undernutrition affects the methylation and expression of the leptin gene in adults: implication for the understanding of metabolic syndrome.* FASEB J, 2011. **25**(9): p. 3271-8.

47. Lesseur, C., et al., *Tissue-specific Leptin promoter DNA methylation is associated with maternal and infant perinatal factors.* Mol Cell Endocrinol, 2013. **381**(1-2): p. 160-7.
48. Lesseur, C., et al., *Maternal obesity and gestational diabetes are associated with placental leptin DNA methylation.* Am J Obstet Gynecol, 2014. **211**(6): p. 654 e1-9.
49. Petersen, P.S., et al., *Deficiency of the GPR39 receptor is associated with obesity and altered adipocyte metabolism.* FASEB J, 2011. **25**(11): p. 3803-14.
50. Janani, C. and B.D. Ranjitha Kumari, *PPAR gamma gene--a review.* Diabetes Metab Syndr, 2015. **9**(1): p. 46-50.
51. Androutsopoulos, V.P., et al., *A mechanistic overview of health associated effects of low levels of organochlorine and organophosphorous pesticides.* Toxicology, 2013. **307**: p. 89-94.
52. Pillai, H.K., et al., *Ligand binding and activation of PPARgamma by Firemaster(R) 550: effects on adipogenesis and osteogenesis in vitro.* Environ Health Perspect, 2014. **122**(11): p. 1225-32.
53. Grimaldi, M., et al., *Reporter Cell Lines for the Characterization of the Interactions between Human Nuclear Receptors and Endocrine Disruptors.* Front Endocrinol (Lausanne), 2015. **6**: p. 62.
54. Khateeb, J., et al., *Paraoxonase 1 (PON1) expression in hepatocytes is upregulated by pomegranate polyphenols: a role for PPAR-gamma pathway.* Atherosclerosis, 2010. **208**(1): p. 119-125.
55. Camps, J., et al., *PPARs in Regulation of Paraoxonases: Control of Oxidative Stress and Inflammation Pathways.* PPAR Res, 2012. **2012**: p. 616371.
56. Khateeb, J., et al., *Urokinase-type plasminogen activator downregulates paraoxonase 1 expression in hepatocytes by stimulating peroxisome proliferator-activated receptor-gamma nuclear export.* Arterioscler Thromb Vasc Biol, 2012. **32**(2): p. 449-58.
57. Divers, J., et al., *Admixture mapping of coronary artery calcified plaque in African Americans with type 2 diabetes mellitus.* Circ Cardiovasc Genet, 2013. **6**(1): p. 97-105.
58. Parks, B.W., et al., *Genetic control of obesity and gut microbiota composition in response to high-fat, high-sucrose diet in mice.* Cell Metab, 2013. **17**(1): p. 141-52.
59. Fox, C.S., et al., *Genome-wide association for abdominal subcutaneous and visceral adipose reveals a novel locus for visceral fat in women.* PLoS Genet, 2012. **8**(5): p. e1002695.
60. Baron, R.M. and D.A. Kenny, *The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations.* J Pers Soc Psychol, 1986. **51**(6): p. 1173-82.
61. Tingley, D., et al., *mediation: R Package for Causal Mediation Analysis.* Journal of Statistical Software, 2014. **59**(5).
62. Pinero, J., et al., *DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes.* Database (Oxford), 2015. **2015**: p. bav028.
63. Verhulst, P.J., et al., *GPR39, a receptor of the ghrelin receptor family, plays a role in the regulation of glucose homeostasis in a mouse model of early onset diet-induced obesity.* J Neuroendocrinol, 2011. **23**(6): p. 490-500.
64. Zhang, J.V., et al., *Obestatin, a peptide encoded by the ghrelin gene, opposes ghrelin's effects on food intake.* Science, 2005. **310**(5750): p. 996-9.
65. Cui, Y., et al., *OPCML is a broad tumor suppressor for multiple carcinomas and lymphomas with frequently epigenetic inactivation.* PLoS One, 2008. **3**(8): p. e2990.
66. Sellar, G.C., et al., *OPCML at 11q25 is epigenetically inactivated and has tumor-suppressor function in epithelial ovarian cancer.* Nat Genet, 2003. **34**(3): p. 337-43.
67. Wu, Y., et al., *Methylation profiling identified novel differentially methylated markers including OPCML and FLRT2 in prostate cancer.* Epigenetics, 2016. **11**(4): p. 247-58.
68. Li, C., et al., *OPCML is frequently methylated in human colorectal cancer and its restored expression reverses EMT via downregulation of smad signaling.* Am J Cancer Res, 2015. **5**(5): p. 1635-48.
69. Volkow, N.D., G.J. Wang, and R.D. Baler, *Reward, dopamine and the control of food intake: implications for obesity.* Trends Cogn Sci, 2011. **15**(1): p. 37-46.
70. Murray, S., et al., *Hormonal and neural mechanisms of food reward, eating behaviour and obesity.* Nat Rev Endocrinol, 2014. **10**(9): p. 540-52.

71. Vucetic, Z., et al., *Maternal high-fat diet alters methylation and gene expression of dopamine and opioid-related genes.* Endocrinology, 2010. **151**(10): p. 4756-64.
72. Paternain, L., et al., *Transcriptomic and epigenetic changes in the hypothalamus are involved in an increased susceptibility to a high-fat-sucrose diet in prenatally stressed female rats*. Neuroendocrinology, 2012. **96**(3): p. 249-60.
73. Ong, Z.Y., J.R. Gugusheff, and B.S. Muhlhausler, *Perinatal overnutrition and the programming of food preferences: pathways and mechanisms.* J Dev Orig Health Dis, 2012. **3**(5): p. 299-308.
74. Palmer, A.A., et al., *Prenatal protein deprivation alters dopamine-mediated behaviors and dopaminergic and glutamatergic receptor binding.* Brain Res, 2008. **1237**: p. 62-74.
75. Vucetic, Z., et al., *Early life protein restriction alters dopamine circuitry.* Neuroscience, 2010. **168**(2): p. 359-70.
76. Wright, T.M., et al., *Exposure to maternal consumption of cafeteria diet during the lactation period programmes feeding behaviour in the rat.* Int J Dev Neurosci, 2011. **29**(8): p. 785-93.
77. Teegarden, S.L., A.N. Scott, and T.L. Bale, *Early life exposure to a high fat diet promotes long-term changes in dietary preferences and central reward signaling.* Neuroscience, 2009. **162**(4): p. 924-32.
78. Torres-Altoro, M.I., et al., *Organophosphates dysregulate dopamine signaling, glutamatergic neurotransmission, and induce neuronal injury markers in striatum.* J Neurochem, 2011. **119**(2): p. 303-13.
79. Slotkin, T.A., *Does early-life exposure to organophosphate insecticides lead to prediabetes and obesity?* Reprod Toxicol, 2011. **31**(3): p. 297-301.
80. Richard, D., Q. Huang, and E. Timofeeva, *The corticotropin-releasing hormone system in the regulation of energy balance in obesity.* Int J Obes Relat Metab Disord, 2000. **24 Suppl 2**: p. S36-9.
81. Richard, D., Q. Lin, and E. Timofeeva, *The corticotropin-releasing factor family of peptides and CRF receptors: their roles in the regulation of energy balance.* Eur J Pharmacol, 2002. **440**(2-3): p. 189-97.
82. Sharma, R. and M.A. Banerji, *Corticotropin releasing factor (CRF) and obesity.* Maturitas, 2012. **72**(1): p. 1-3.
83. Uehara, Y., et al., *Hypothalamic corticotropin-releasing hormone is a mediator of the anorexigenic effect of leptin.* Diabetes, 1998. **47**(6): p. 890-3.
84. Sutton, A.K., et al., *Control of food intake and energy expenditure by Nos1 neurons of the paraventricular hypothalamus.* J Neurosci, 2014. **34**(46): p. 15306-18.
85. Leshan, R.L., et al., *Leptin action through hypothalamic nitric oxide synthase-1-expressing neurons controls energy balance.* Nat Med, 2012. **18**(5): p. 820-3.
86. Sansbury, B.E. and B.G. Hill, *Regulation of obesity and insulin resistance by nitric oxide.* Free Radic Biol Med, 2014. **73**: p. 383-99.
87. Naseh, M., et al., *Alterations in nitric oxide synthase-expressing neurons in the forebrain regions of rats after developmental exposure to organophosphates.* Neurotoxicol Teratol, 2013. **37**: p. 23-32.
88. Naseh, M. and J. Vatanparast, *Enhanced expression of hypothalamic nitric oxide synthase in rats developmentally exposed to organophosphates.* Brain Res, 2014. **1579**: p. 10-9.
89. Snodgrass-Belt, P., J.L. Gilbert, and F.C. Davis, *Central administration of transforming growth factor-alpha and neuregulin-1 suppress active behaviors and cause weight loss in hamsters.* Brain Res, 2005. **1038**(2): p. 171-82.
90. Ennequin, G., et al., *Neuregulin 1 affects leptin levels, food intake and weight gain in normal-weight, but not obese, db/db mice.* Diabetes Metab, 2015. **41**(2): p. 168-72.
91. Wu, G., et al., *Mechanism and clinical evidence of lipocalin-2 and adipocyte fatty acid-binding protein linking obesity and atherosclerosis.* Diabetes Metab Res Rev, 2014. **30**(6): p. 447-56.
92. Hotamisligil, G.S. and D.A. Bernlohr, *Metabolic functions of FABPs--mechanisms and therapeutic implications.* Nat Rev Endocrinol, 2015. **11**(10): p. 592-605.
93. Furuhashi, M., et al., *Fatty Acid-Binding Protein 4 (FABP4): Pathophysiological Insights and Potent Clinical Biomarker of Metabolic and Cardiovascular Diseases.* Clin Med Insights Cardiol, 2014. **8**(Suppl 3): p. 23-33.

94. Jiang, Y., et al., *FABP4-mediated homocysteine-induced cholesterol accumulation in THP-1 monocyte-derived macrophages and the potential epigenetic mechanism.* Mol Med Rep, 2016.
95. Yang, A.N., et al., *High-methionine diets accelerate atherosclerosis by HHcy-mediated FABP4 gene demethylation pathway via DNMT1 in ApoE(-/-) mice.* FEBS Lett, 2015. **589**(24 Pt B): p. 3998-4009.
96. Marrades, M.P., et al., *Orchestrated downregulation of genes involved in oxidative metabolic pathways in obese vs. lean high-fat young male consumers.* J Physiol Biochem, 2011. **67**(1): p. 15-26.
97. Mussig, K., et al., *Genetic variants in MTNR1B affecting insulin secretion.* Ann Med, 2010. **42**(6): p. 387-93.
98. Nagorny, C. and V. Lyssenko, *Tired of diabetes genetics? Circadian rhythms and diabetes: the MTNR1B story?* Curr Diab Rep, 2012. **12**(6): p. 667-72.
99. Karamitri, A., et al., *Minireview: Toward the establishment of a link between melatonin and glucose homeostasis: association of melatonin MT2 receptor variants with type 2 diabetes.* Mol Endocrinol, 2013. **27**(8): p. 1217-33.
100. Liu, R., et al., *Correlation of functional GRIN2A gene promoter polymorphisms with schizophrenia and serum D-serine levels.* Gene, 2015. **568**(1): p. 25-30.
101. Endele, S., et al., *Mutations in GRIN2A and GRIN2B encoding regulatory subunits of NMDA receptors cause variable neurodevelopmental phenotypes.* Nat Genet, 2010. **42**(11): p. 1021-6.
102. Yoo, H.J., et al., *Family based association of GRIN2A and GRIN2B with Korean autism spectrum disorders.* Neurosci Lett, 2012. **512**(2): p. 89-93.
103. Carvill, G.L., et al., *GRIN2A mutations cause epilepsy-aphasia spectrum disorders.* Nat Genet, 2013. **45**(9): p. 1073-6.
104. Lal, D., et al., *Investigation of GRIN2A in common epilepsy phenotypes.* Epilepsy Res, 2015. **115**: p. 95-9.
105. Ninomiya, M., et al., *Cortical neurons from intrauterine growth retardation rats exhibit lower response to neurotrophin BDNF.* Neurosci Lett, 2010. **476**(2): p. 104-9.
106. Wang, L., et al., *Polymorphism in maternal LRP8 gene is associated with fetal growth.* Am J Hum Genet, 2006. **78**(5): p. 770-7.
107. Shen, G.Q., L. Li, and Q.K. Wang, *Genetic variant R952Q in LRP8 is associated with increased plasma triglyceride levels in patients with early-onset CAD and MI.* Ann Hum Genet, 2012. **76**(3): p. 193-9.
108. Shen, G.Q., et al., *Multi-allelic haplotype association identifies novel information different from single-SNP analysis: a new protective haplotype in the LRP8 gene is against familial and early-onset CAD and MI.* Gene, 2013. **521**(1): p. 78-81.
109. Shen, G.Q., et al., *A novel molecular diagnostic marker for familial and early-onset coronary artery disease and myocardial infarction in the LRP8 gene.* Circ Cardiovasc Genet, 2014. **7**(4): p. 514-20.
110. Guo, T., et al., *Integrative mutation, haplotype and G x G interaction evidence connects ABGL4, LRP8 and PCSK9 genes to cardiometabolic risk.* Sci Rep, 2016. **6**: p. 37375.
111. Alegria-Torres, J.A., et al., *Q192R polymorphism of paraoxonase 1 gene associated with insulin resistance in Mexican children.* Arch Med Res, 2015. **46**(1): p. 78-83.
112. Huen, K., et al., *Associations of PON1 and genetic ancestry with obesity in early childhood.* PLoS One, 2013. **8**(5): p. e62565.
113. Huen, K., et al., *PON1 as a model for integration of genetic, epigenetic, and expression data on candidate susceptibility genes.* Environ Epigenet, 2015. **1**(1).
114. Coombes, R.H., et al., *Human paraoxonase 1 hydrolysis of nanomolar chlorpyrifos-oxon concentrations is unaffected by phenotype or Q192R genotype.* Toxicol Lett, 2014. **230**(1): p. 57-61.
115. Cole, T.B., et al., *Expression of human paraoxonase (PON1) during development.* Pharmacogenetics, 2003. **13**(6): p. 357-364.
116. Costa, L.G., et al., *Paraoxonase 1 (PON1) as a genetic determinant of susceptibility to organophosphate toxicity.* Toxicology, 2013. **307**: p. 115-22.

117. Huang, Y.T., et al., *Epigenome-wide profiling of DNA methylation in paired samples of adipose tissue and blood.* Epigenetics, 2016. **11**(3): p. 227-36.

Supplementary Information

Supplementary tables 1 to 6 can be found in following dropbox folder:

https://www.dropbox.com/sh/d9lq9mipqkoezx3/AAC9RCXgWS7P6DS71ybJqWNJa?dl=0

Supplementary Table 1: Primer Sequences.

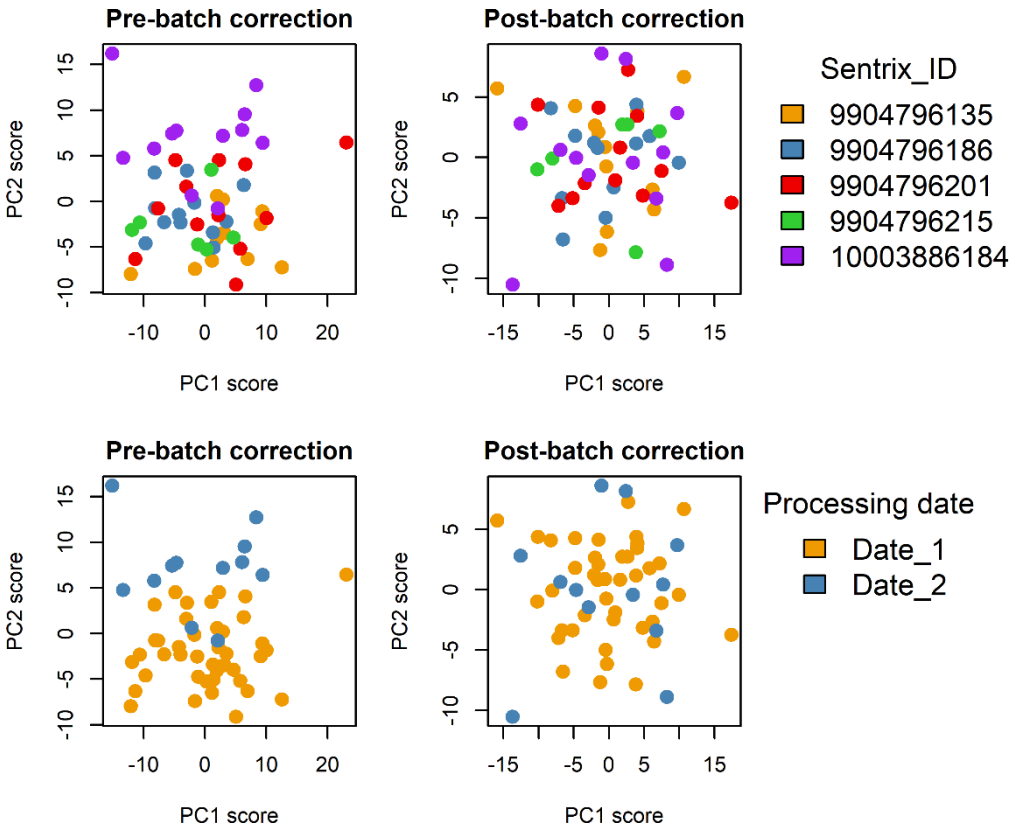
Supplementary Table 2: Sig-DMPs overlapping with DMRs (interaction model).

Supplementary Table 3: Outcome DMR pyrosequencing.

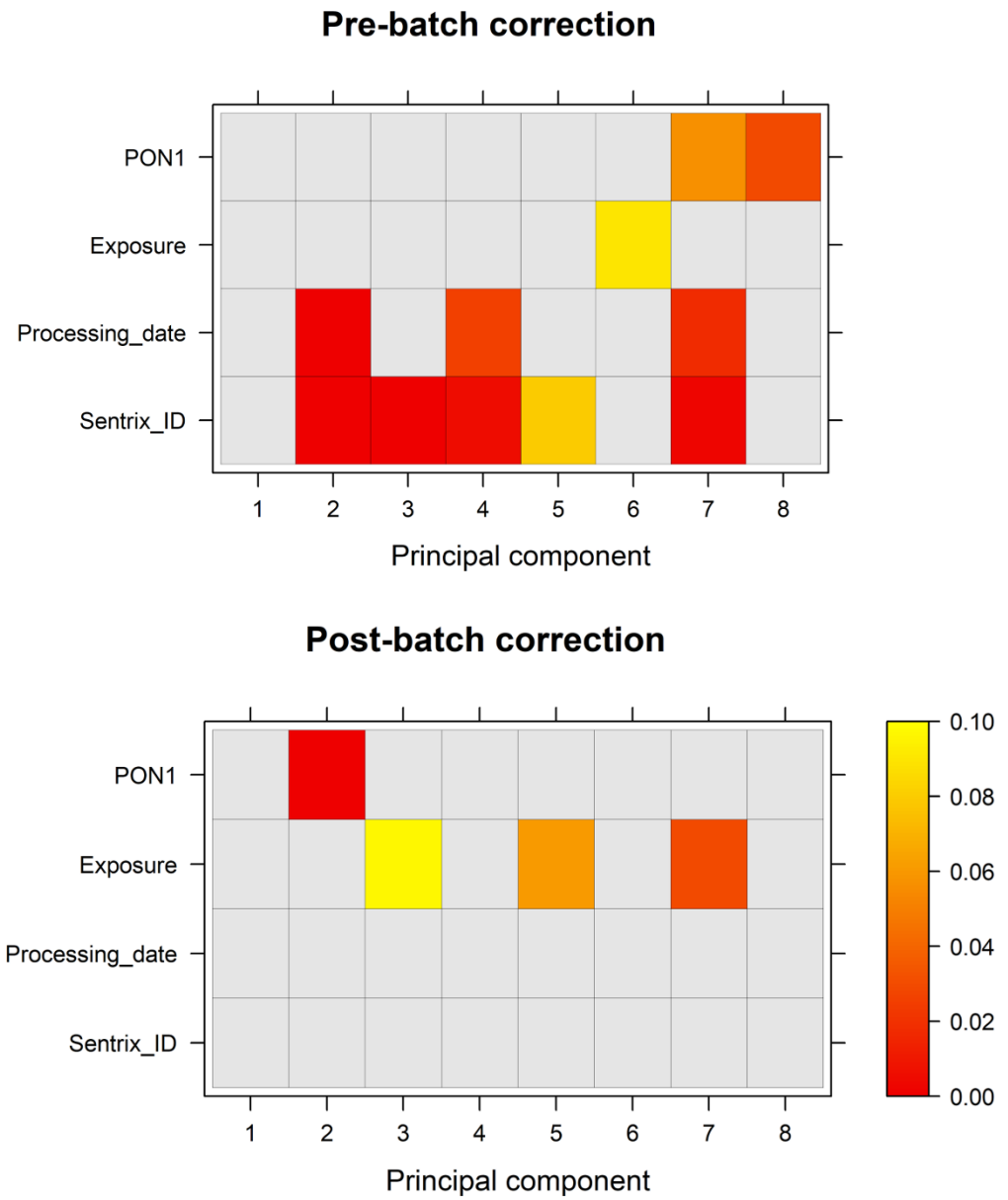
Supplementary Table 4: Enrichment of TFBS for DMPs significant in the interaction model.

Supplementary Table 5: Outcome statistics and gene disease associations of (partial) mediators between pesticide exposure and bodyfat measures in PON1-QR allele carriers.

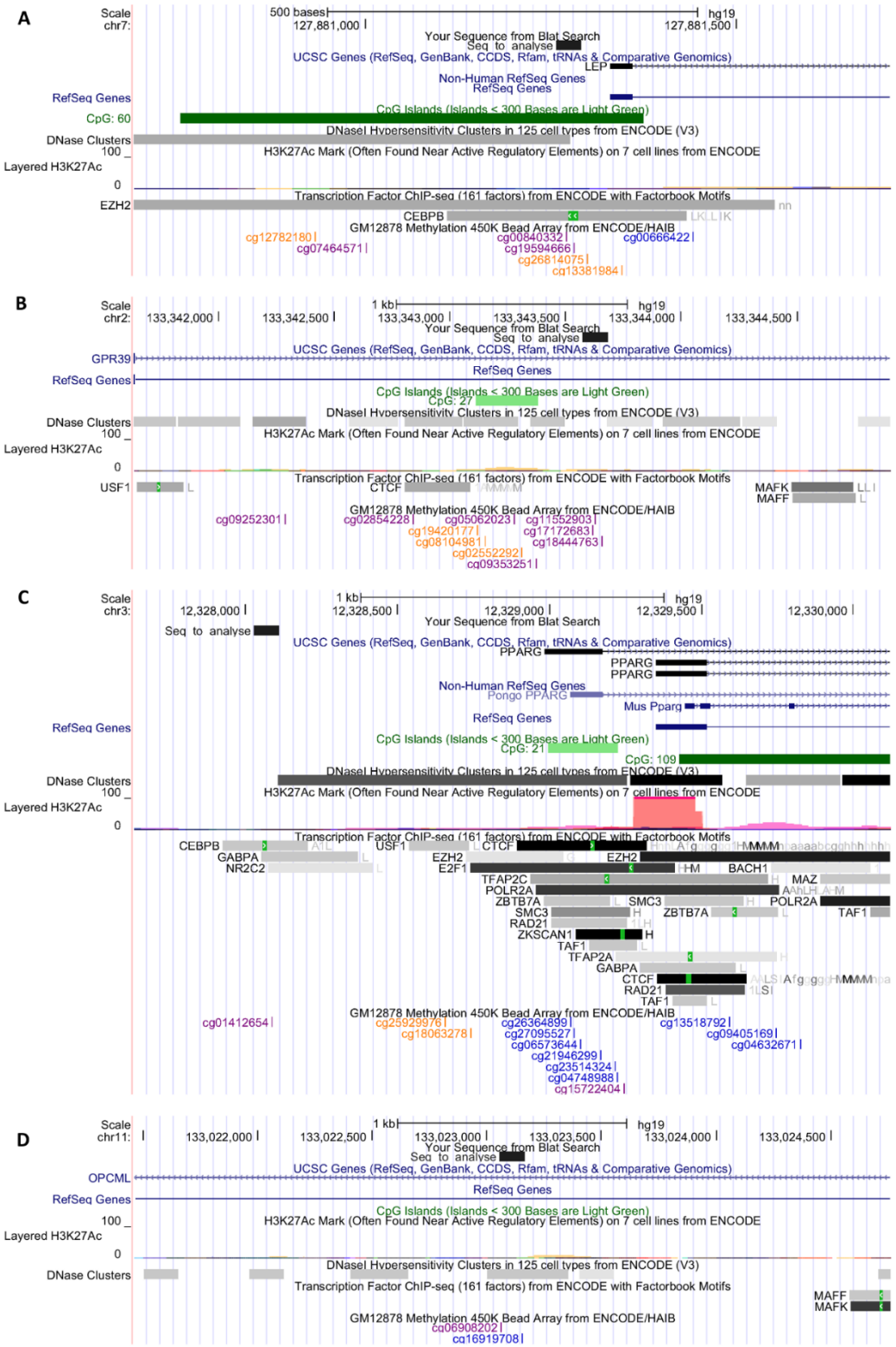
Supplementary Table 6: Association between estimated blood cell counts and health outcomes.



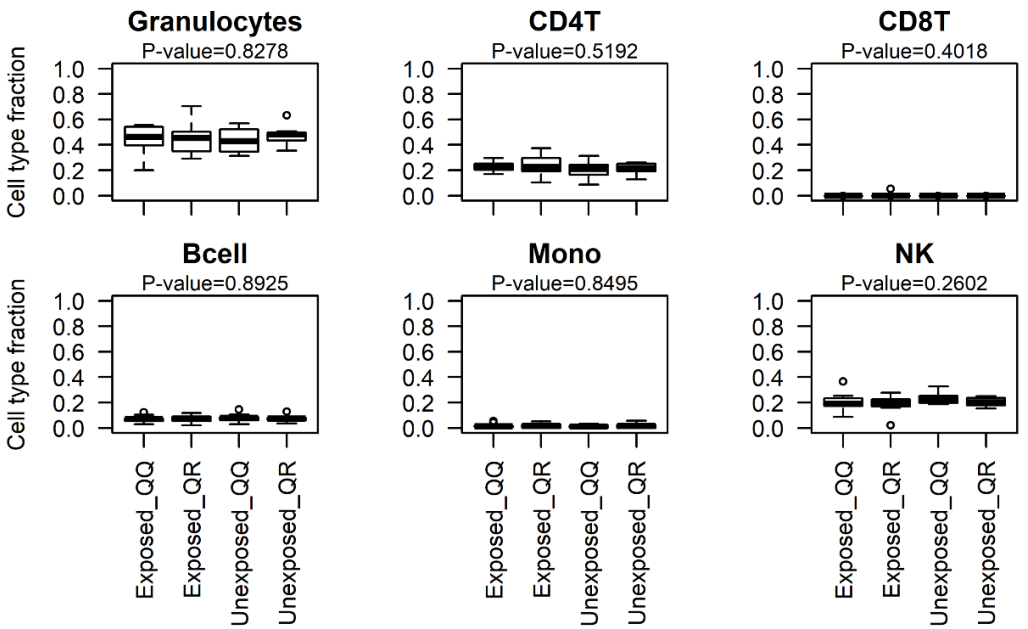
Supplementary Figure 1: PCA before and after batch effect correction for Sentrix_ID and processing date using ComBat.



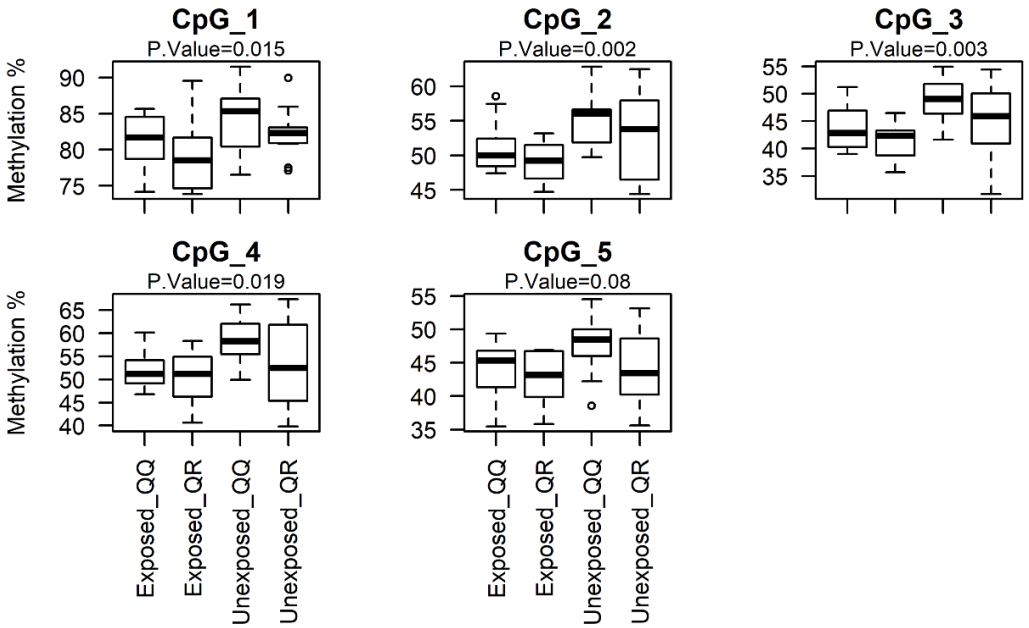
Supplementary Figure 2: Associations between the first eight principal components and covariates before and after ComBat batch correction. Associations between principal components and Sentrix_ID were measured using the Kruskal-Wallis test. Associations between principal components and processing date, exposure and *PON1* Q192R genotype were measured using the two-sided Wilcoxon sum rank test.


A
Scale
chr7:
500 bases
127,881,000
127,881,500
hg19
Your Sequence from Blat Search
Seq to analyse
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
LEP
Non-Human RefSeq Genes
RefSeq Genes
RefSeq Genes
CpG Islands (Islands < 300 Bases are Light Green)
CpG: 60
DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE (V3)
DNase Clusters
100
H3K27Ac Mark (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE
Layered H3K27Ac
0
Transcription Factor ChIP-seq (161 factors) from ENCODE with Factorbook Motifs
EZH2
CEBPB
GM12878 Methylation 450K Bead Array from ENCODE/HAIB
cg12782180
cg07464571
cg00840332
cg19594666
cg26814075
cg13381984
cg00666422
B
Scale
chr2:
1 kb
133,342,000
133,342,500
133,343,000
133,343,500
133,344,000
133,344,500
hg19
Your Sequence from Blat Search
Seq to analyse
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
GPR39
RefSeq Genes
RefSeq Genes
CpG Islands (Islands < 300 Bases are Light Green)
CpG: 27
DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE (V3)
DNase Clusters
100
H3K27Ac Mark (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE
Layered H3K27Ac
0
Transcription Factor ChIP-seq (161 factors) from ENCODE with Factorbook Motifs
USF1
CTCF
MAFK
MAFF
GM12878 Methylation 450K Bead Array from ENCODE/HAIB
cg09252301
cg02854228
cg05062023
cg11552903
cg19420177
cg17172683
cg08104981
cg18444763
cg02552292
cg09353251
C
Scale
chr3:
1 kb
12,328,000
12,328,500
12,329,000
12,329,500
12,330,000
hg19
Your Sequence from Blat Search
Seq to analyse
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
PPARG
Non-Human RefSeq Genes
Pongo PPARG
Mus Pparg
RefSeq Genes
RefSeq Genes
CpG Islands (Islands < 300 Bases are Light Green)
CpG: 21
CpG: 109
DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE (V3)
DNase Clusters
100
H3K27Ac Mark (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE
Layered H3K27Ac
0
Transcription Factor ChIP-seq (161 factors) from ENCODE with Factorbook Motifs
CEBPB
GABPA
NR2C2
USF1
CTCF
EZH2
E2F1
BACH1
TFAP2C
MAZ
POLR2A
ZBTB7A
SMC3
RAD21
ZKSCAN1
TAF1
TFAP2A
GABPA
CTCF
RAD21
GM12878 Methylation 450K Bead Array from ENCODE/HAIB
cg01412654
cg25929976
cg18063278
cg26364899
cg27095527
cg06573644
cg21946299
cg23514324
cg04748988
cg15722404
cg13518792
cg09405169
cg04632671
D
Scale
chr11:
1 kb
133,022,000
133,022,500
133,023,000
133,023,500
133,024,000
133,024,500
hg19
Your Sequence from Blat Search
Seq to analyse
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
OPCML
RefSeq Genes
RefSeq Genes
CpG Islands (Islands < 300 Bases are Light Green)
100
H3K27Ac Mark (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE
Layered H3K27Ac
0
DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE (V3)
DNase Clusters
Transcription Factor ChIP-seq (161 factors) from ENCODE with Factorbook Motifs
MAFF
MAFK
GM12878 Methylation 450K Bead Array from ENCODE/HAIB
cg06908202
cg16919708

Supplementary Figure 3: Genomic location of the pyrosequencing assays represented as a UCSC genome browser track. The first track indicates the sequence analysed by pyrosequencing (Seq_to_analyse). Other custom tracks include: CGIs, Dnase I hypersensitivity clusters, H3K27ac histone marks, TFBSs and the Illumina 450K methylation probes. **A)** *LEP* assay **B)** *GPR39* assay **C)** *PPARG* assay and **D)** *OPCML* assay.



Supplementary Figure 4: Relative cell type contribution estimated by the Houseman approach. Differences in cell type composition between the exposure groups were measured using one-way ANOVA.



Supplementary Figure 5: Outcome of GPR39 DMR pyrosequencing. Boxplots showing methylation differences between the exposure groups in the *GPR39* pyrosequencing region. P-values shown are those of the exposure effect.

4

Identification of CALD1 DNA Methylation as a saliva specific biomarker for early life psychosocial deprivation stress associated with institutional care

Stacy Drury[1], Ken Declerck[2], Wim Vanden Berghe[2], Patrick De Boever[3]

[1] The Brain Institute, Tulane University, New Orleans, LA, USA.
[2] Laboratory of Protein Chemistry, Proteomics and Epigenetic Signaling, Department of Biomedical Sciences, University of Antwerp, Wilrijk, Belgium.
[3] Environmental Risk and Health Unit, Flemish Institute for Technological Research (VITO), Mol, Belgium.

Manuscript in preparation.

Contributions: Ken Declerck analyzed the data, helped in interpreting the results and wrote results section, part of materials & methods and discussion.

Abstract |Epigenetic DNA methylation changes are hypothesized to be one mechanism through which early life stress experiences shape adverse neurodevelopment across social, emotional, behavioral, physiological and neurobiological health outcomes associated with psychosocial deprivation during institutional care. In this study we characterized genes exhibiting differential methylation and gene expression changes in saliva, associated with changes in cortisol levels during the Trier Social Stress Test (TSTT) in children enrolled in the Bucharest Early Intervention Project (BEIP). The latter is a unique randomized controlled trial of foster care compared to care as usual for children abandoned and placed in institutions at birth or early in life. We identified a set of 30 DMRs associated genes which are involved in nervous system and neuronal development, related to cognition, behavior and learning and psychological disorders. Further comparison with gene expression profiles before and after the TSTT identified a nuclear receptor coordinated stress dependent gene network, involving the glucocorticoid receptor. One of the genes with the highest correlation between TSST specific changes in cortisol levels, DNA methylation and gene expression was *CALD1*, a gene with established links to glucocorticoid stress responsiveness, neuronal migration and also the impact of cortisol stress hormones on neuronal morphology, dendritic spines and migration. Furthermore, chronic stress-induced shrinkage of dendritic spine subtypes has been shown to impair feedback mechanisms of the HPA-axis, cause deterioration of sensory functions and attrition of neuroplasticity. These findings revealed an important mechanistic link for CALD1 in the lasting neurodevelopmental impacts of early life stress. Our results offer an exciting glimpse into the potential utility of salivary diagnostics to define the underlying mechanisms through which early life adversity is biologically embedded and creates a unique opportunity to evaluate the impact of interventions on biological processes without invasive protocols and with methodology that can be utilized in remote locations and processed at later time points.

Introduction

Early severe psychosocial deprivation associated with institutional care has lasting negative impacts across social, emotional, behavioral, physiological, neurobiological and health outcomes. Substantial evidence indicates that providing children stable and sensitive caregiving can result in significant recovery, albeit at times partial and unequal across domains. While providing enhanced caregiving results in significant improvements not all children improve equally and the timing of when the change in caregiving occurs further influences long-term outcomes [1, 2]. Despite a substantial

body of research on the consequences of institutional care, and the benefits of foster care to mitigate lasting impacts, little is known about the molecular mechanisms that contribute to both the initial negative impacts and the positive cross-domain effects of improved caregiving [3, 4].

Epigenetic changes, particularly methylation, are hypothesized to be one mechanism through which early life experiences, both negative and positive, influence development [5, 6]. Methylation refers to the addition of a methyl group to the DNA strand leading to conformational changes to the DNA structure and/or alterations to the accessibility of the DNA sequence to transcriptional machinery that converts the DNA sequence into RNA. The majority of studies exploring the epigenetic consequences of early adversity focus on the presence of methylation in specific regulatory regions of a gene, termed CGIs, where increased methylation predominately, but not exclusively, leads to decreased transcription. These changes in gene expression and subsequent differences in protein production are hypothesized to drive observed physiologic and developmental differences that result in the observed phenotypes associated with early adversity including changes to the regulation of the hypothalamic pituitary axis (HPA), increased risk of psychopathology, and negative health trajectories. However, to date, few studies have demonstrated the direct link between exposure to early adversity, alterations in methylation, differences in gene expression and their downstream contribution to identified phenotypes, representing a significant gap in our ability to mechanistically define how early adversity is biologically embedded.

Preclinical rodent and non-human primate (NHP) studies have reported both gene specific and whole genome alteration in the methylome in association with early adverse caregiving [7-10]. More recent studies have reported similar alterations in gene specific methylation across rodent, NHP and humans suggesting that, in some cases, the epigenetic pathways altered by deviations from species expected caregiving are evolutionarily conserved [11]. To date five studies have examined DNA methylation differences specifically in association with institutional rearing, four exploring changes across the genome and one focused on methylation differences in stress related genes [12-16]. In the first small study, children with a variable history of institutional care were compared to children reared in typical homes. Small, but significant, methylation differences, in DNA extracted from whole blood, were found at 914 sites across the whole genome. Notably while the majority of these sites were hyper-methylated in children with a history of institutional care, a subset of sites were hypo-methylated. The very small sample size and heterogeneous nature of the previously institutionalized children suggests that replication is needed [12]. A second study comparing children internationally adopted from institutions in Eastern Europe or Russia to children matched for family income and education reared in their biological families also

reported methylation differences after removing invariant sites and correcting for cell type differences. Significant differences were limited to 30 sites in 19 genes, none of which were in the previously reported study and none that were linked directly to clinically significant outcomes [15]. Notably significant differences in the CD4T/CD8T cell ratios between institutionalized child and controls were found in both these participants and in a re-analysis of the subjects in the Naumova study. The authors noted that differences in cell types likely account for a substantial amount of differential methylation in studies that fail to account for this variability. This finding, notable on its own for the implications for inflammatory function in children with a history of institutional care, also highlights one recurrent challenge- potential differential methylation across cell types. A third whole methylome study, in buccal DNA, explored the link between methylation and the duration of institutional rearing, a factor known to substantially influence short and long-term outcomes [14]. Although again in a limited sample size (n=32 total), they did identify alterations across 9 CpG sites in CYP2E1, an abundantly expressed and multifunctional member of the cytochrome p450 super family. These altered methylation sites were not only different based on duration of institutional care but were also correlated with cognitive function and theory of mind, two common areas of lasting impact following institutional exposure. How altered methylation in this gene contributed to lasting negative neurodevelopmental outcomes remains uncertain although preclinical animal models suggest that stress exposure and decreased CYP2E1 levels may influence outcomes through serotonergic pathways [17]. A very recent genome-wide methylation study in infants (8-35 months of age) reared in institutions, detected 164 differentially methylated sites in blood compared to infants reared in biological families [16]. Furthermore, some of the epigenetic markers were also associated with adaptive behavior skills indicating a possible link between institutional care and behavioral deficits.

Although whole genome studies have advantages, challenges facing studies in post-institutionalized populations, specifically limited sample size and variability in both the duration of exposure to institutional care and subsequent caregiving after placement in foster or adopted families, indicate that alternative, ideally complimentary, approaches are warranted. One approach is to explore specific methylation changes in genes functionally related to aberrant outcomes in more homogenous post-institutionalized populations, an approach utilized in a study by *Non et al.* [13]. The Bucharest Early Intervention Project (BEIP) is the only randomized controlled trial of foster care compared to care as usual for children abandoned and placed in institutions at birth or early in life. This prospective longitudinal study began when children were, on average, 22 months of age. Children were recruited from institutions in Bucharest

and randomized to a newly created foster care (FCG) or to continued institutional care (care as usual, CAUG). A group of children never institutionalized children (NIG) were recruited from the same maternity hospitals at baseline [18] and area schools at the age 8 [19]. In this study, DNA methylation in functional regions of *SERT* and *FKBP5* were significantly correlated with the duration of institutional care where both genes demonstrated negative correlations between the percentage of time a child spent in institutional care and methylation [13]. Notably while again the number of individuals was limited (n=127), the sample size was significantly larger than the previous whole genome studies and targeted two genes with established links to cortisol reactivity, a phenotype that exhibited persistent group differences at age 12 in this cohort [20]. Despite this targeted approach methylation differences did not significantly correlate with group differences in cortisol reactivity, most likely because the genetic architecture of the acute cortisol response is multi-genetic in etiology with no individual gene contributing a large portion of the effect size. Although informative, these studies have been unable to directly link altered methylation with specific outcomes and none have examined associations with gene expression, the presumed downstream functional consequence of altered methylation.

Methylation differences are hypothesized to influence outcomes through their down-stream impact on gene expression and function. In preclinical animal models the ability to concurrently measure gene expression and methylation in target tissues (e.g. the brain) permits the direct examination of the correlation between altered methylation, gene expression and specific outcomes, an approach unavailable for the majority of human studies. Beyond the challenge of tissue accessibility, the potential differential patterns found across tissues creates another analytic issue as often the targeted phenotype cannot be concurrently measured within that same sample [21]. An ideal biological source would be easily obtainable in large volumes, contain DNA, RNA and protein products directly linked to a testable phenotype. Saliva represents a unique and under-utilized biospecimen for biological studies that meets these criteria. First, saliva is accessible unobtrusively from a wide range of participants, including vulnerable or at risk populations, acknowledging that collection methods may differ. Second, saliva can be replenished, thereby permitting repeated measurements across short and long periods of time. Third, salivary mRNA expression analysis has recently been established and evidence exists that, in some instances mRNA expression studies in saliva may be more applicable than studies in peripheral blood [21, 22]. Fourth, multiple analytes including immune markers (e.g salivary IgA, salivary alpha amylase) as well as endocrine markers (cortisol, testosterone, DHEA) are readily measurable in saliva and highly correlated with levels in peripheral blood. These findings suggest that one approach to establishing molecular mechanistic pathways would be to examine, within saliva,

concurrent DNA methylation, gene expression and protein levels. One marker with extensive validation in saliva that has been linked to altered methylation in several genes and exhibits both an acute response and diurnal variation impacted by early institutional rearing is cortisol.

Several previous studies have reported associations between DNA methylation and cortisol levels [23]. For example, methylation levels in DNA from peripheral blood in *spindle and kinetochore associated protein 2* (*SKA2*) gene were significantly associated with cortisol reactivity during a social stressor and also mediated the relationship between cortisol levels and childhood trauma exposure [24]. An additional study found that methylation in FKBP5 moderated the association between resistant attachment and cortisol reactivity in a group of 298 14-month-old infants [25]. Despite established evidence of the utility of the salivary RNA expressome to date no studies have examined concurrently salivary methylation, gene expression and cortisol. One previous study in adults revealed an association between cortisol response to a social stressor, and DNA methylation and gene expression of the glucocorticoid receptor (GR) gene in peripheral blood in individuals with post-traumatic stress disorder (PTSD) compared to healthy controls [26]. Unfortunately, no explanation for why cortisol was not also measured in blood was provided and the samples were not temporally collected limiting the ability to draw causal relationships despite efforts to define alterations at the level of DNA methylation and expression with cortisol levels.

The number of studies utilizing salivary RNA are increasing, however, caveats both to the within individual variation and the integrity of RNA species remain [27]. One approach to minimize the confounding of random variation in expression is to measure RNA expression concurrently with a known protein product and ideally utilize repeated measures of the protein in association with RNA expression. Samples taken during paradigms that are expected to produce differences in physiologic processes would provide construct validity if the expression of genes involved in the regulation of that targeted physiologic processes were associated. The Trier Social Stressor Test (TSST) is a social evaluative threat paradigm that results in individual differences in cortisol expression that have been measured in both peripheral blood and saliva and reflect the reactivity of the hypothalamic pituitary axis (HPA) in response to stress [28, 29]. In children exposed to early institutional care significant differences in cortisol reactivity elicited in response to this paradigm exist [20]. Studies have shown high correlation between cortisol levels measured in peripheral blood and saliva, suggesting that changes in salivary composition are reflective of the physiological changes associated with activation of the HPA axis [29]. Studies examining the ability of salivary gene expression to predict group differences in cortisol levels and whether gene expression also was associated with group differences in methylation levels represents an

unprecedented opportunity to test both mechanism and the functional significance of methylation changes.

As an important proof of concept and to take a needed next step toward defining mechanistic changes in gene expression that contribute to persistent physiologic differences following institutional rearing this study examined the association between exposure to institutional caregiving and both salivary DNA methylation and gene expression. Further, this study tested whether genes exhibiting differential methylation that were also correlated with gene expression levels demonstrated an association with cortisol levels collected concurrently during the TSST in children enrolled in the BEIP at age 12. The randomized prospective study design provides a unique opportunity for the exploration of the association between a well-characterized exposure to severe psychosocial deprivation and differential methylation and gene expression during a defined social stressor in saliva and critically how these differences related to concurrent physiologic changes in cortisol levels.

Materials and methods

Subjects

Participants were a subgroup of children enrolled in the BEIP [18], a longitudinal randomized controlled trial of foster care compared to care as usual for children in Romanian institutions, described in detail elsewhere and conducted entirely in Romania [1, 30]. 136 children, between 6 and 31 months of age, residing in six institutions in Romania were initially enrolled (ever institutionalized, EIG), and following baseline assessments, randomly assigned to CAUG (*n*=68) or FCG (*n*=68). The foster care system utilized in this study was created specifically for this project as an intentional alternative to institutional care. The related ethical considerations have been described in detail [31, 32]. A reference group of children without any history of institutional rearing were recruited either from the same maternity hospitals in which the EIG were born or, for later recruitment, from the same schools. As mandated by Romanian law, the Commission on Child Protection provided informed consent for each of the child participants. Following randomization and placement of children in foster care, all subsequent decisions regarding placement were made by the child protection commissions in Romania. The Institutional Review Boards of Children's Hospital of Boston, University of Maryland, and Tulane University approved this study.

This study was based on data obtained from a subset of individuals selected initially for IQ (>60 at age 12) and completeness of data including completion of the entire TSST,

all time points of cortisol and saliva samples available and collection of RNA samples. A second level of selection occurred based on quality of RNA with participants who had a RIN value of less than 4 for either sample excluded due to concerns about quality control. RNA was collected and assayed on 29 individuals. Results of 4 individuals were excluded because of low RIN values and/or failed quality control on either the methylation or the expression data. No differences in percent time institutionalization, IQ or cortisol levels at any time point were found between those subjects included in this analysis and the larger study group. No significant differences in relevant variables were identified for subjects with failed analyses on either array or low-quality RIN and those included in the final analyses.

Social Stressor Paradigm

At age 12 individuals in the BEIP completed the TSST as well as a series of other tasks (**Supplementary Figure 1**) [20]. Participants arrived at the research facility in Bucharest in the afternoon around 1pm. Approximately 30 minutes after arrival, saliva samples were collected by passive drool for cortisol (cort 1) and DNA into a cryovial and RNA (RNA 1) using Oragene salivary RNA kits. Participants then completed three laboratory based procedures designed to elicit physiologic reactivity: (1) a passive peer evaluation task (2) a social evaluative threat task (TSST) a commonly used social stress paradigm known to induce cortisol reactivity and (3) a nonsocial frustration task that required active participation. Each set of tasks was followed by a five-minute recovery. Cortisol and RNA samples were collected at time points presented in **supplementary figure 1**. All samples were collected by passive drool through a plastic straw. For cortisol analyses all samples were frozen immediately at -20 degrees and shipped on dry ice to a laboratory in Boston where they were assayed using commercially available luminescence immunoassay kits (CLIA; IBL). Cortisol assay sensitivities and procedures are described more fully in a previous publication [20]. Salivary RNA sample were collected into Oragene RNA kits and frozen. Samples were also shipped on dry ice to the laboratory of Dr. Drury for subsequent analyses. DNA was extracted directly from the same cryovials as utilized for salivary cortisol levels.

RNA extraction

Saliva was collected using Oragene RNA self-collection kits (RE-100, DNA Genotek, Ontario, Canada) concurrent with expected cortisol baseline (Cort 1/RNA 1) and at expected peak cortisol following the TSST (Cort 3, RNA 2). RNA sample collection frequency was limited by cost. The two-time points were specifically chosen to capture baseline and peak cortisol levels in an effort to permit examination of fold change in

salivary RNA expression between baseline and expected peak with concurrent changes in cortisol levels. RNA was extracted from saliva according to the manufacturer's protocol (DNA Genotek, Ontario, Canada). RNA was then assessed for integrity and purity using QuBit, Nanodrop and Agilent Bioanalyzer to determine RIN.

DNA extraction

Genomic DNA (gDNA) was extracted from 2 ml saliva using the Oragene OG-500 PrepIT kit (DNA Genotek, Ottawa, Canada) according to the manufacturer's instructions. About 500 ng of gDNA was bisulfite converted using the EZ DNA methylation kit (Zymo Research, Cambridge Bioscience, Cambridge, UK) according to the manufacturer's instructions. Bisulfite conversion and quality control was performed as previously described [33].

Genome-wide DNA methylation

DNA methylation profiles were generated with Infinium HumanMethylation450 BeadChip Array (Illumina, San Diego, CA, USA). 4 µl of bisulfite-converted DNA (~150 ng) was used for the whole genome amplification reaction, enzymatic fragmentation, precipitation and resuspension in hybridization buffer. Subsequent steps of DNA methylation analysis were carried out according to the standard Infinium HD Assay Methylation Protocol Guide (Part #15019519, Illumina). The BeadChip images were captured using the Illumina iScan. The raw methylation intensities for each probe were represented as methylation β-values (ranging from 0, unmethylated, to 1, fully methylated) and extracted from GenomeStudio Methylation Module software without background correction and normalization.

Genome-wide gene expression

Total RNA was amplified and labelled to generate complementary RNA (cRNA) using the Quick Amp Labelling (two color) kit (Agilent Technologies) according to the manufacturer's instructions. The single-stranded, labelled cRNA was purified with Qiagen's RNeasy mini spin columns. Yield (at least 825 ng/sample) and specific activity were determined using a NanoDrop Spectrophotometer (NanoDrop Technologies). Cy3 and Cy5 labelled samples showed a specific activity of >8 pmol/µg cRNA, above which the labelling reaction is considered successful by the manufacturer. Equal amounts (825 ng) of both samples were combined and competitively hybridized on 4x44K Whole Human Genome microarray slides (design 014850, Agilent Technologies) for 17 hours using the automated HS4800^{TM} pro hybridization station (Tecan, Männedorf,

Switzerland) according to the manufacturer's instructions. The arrays were scanned on an Agilent DNA microarray scanner (G2565BA) and processed using Agilent Feature Extraction Software (Version 10.7).

DNA methylation analysis

Raw data were imported into R environment using the RnBeads R package [34]. The R package MethylAid was used to identify bad quality samples, making use of the quality control probes present on the methylation array [35]. The default thresholds were used to define outlying samples. Outlying samples were removed for further analysis (n=5), leaving valid data of 22 subjects for further statistical analysis. RnBeads was further used to preprocess the Illumina 450K methylation data [34]. First, Illumina probes were filtered based on following criteria: probes containing a SNP within 3 bp of the analyzed CpG site, bad quality probes based on an iterative greedy cut algorithm using a detection p-value of 0.01 as a threshold for an unreliable measurement, and probes with missing values in at least one sample. Beta-values (ratio of methylated probe intensity versus total probe intensity) were subsequently within-array normalized using the BMIQ method [36]. Next, Illumina probes measuring methylation not at CpG sites and/or located at sex chromosomes were removed. Beta-values were transformed to M-values ($M = \log2(\beta/(1-\beta))$) prior to further analyses. PCA was conducted to detect possible batch effects. Associations between the first eight principal components and possible batch effect covariates (sentrix ID and sentrix position) were measured using the Kruskall-Wallis test or the two-sided Wilcoxon sum rank test. Because we couldn't find any significant association between the principal component scores and the covariates, we decided to not take batch into account in further analyses. For each sample the relative cell type contribution was measured using the reference-based approach described by Houseman and coworkers [37]. Reference methylomes of each blood cell type (granulocyte, CD4+ T-cell, CD8+ T-cell, B-cell, monocyte, NK-cell) were obtained from the study of Reinius et al. using the FlowSorted.Blood.450k R package [38]. Buccal epithelial reference methylomes were retrieved from the GEO dataset GSE48472. The analysis was limited to the 100,000 most variable sites. The top 500 cg-probes associated with the cell types were used to estimate the relative cell type composition in each sample. Differences in cell type contributions between the groups were calculated using the one-way ANOVA test. Finally, differentially methylated probes (DMPs) were detected using the moderated t-test of the limma R package [39]. Gender was incorporated as a covariate in the linear model. P-values were corrected for multiple testing using the method of Hochberg and Benjamini. DMRs were

identified using the DMRcate R package [40]. Given sample size limitations the p-value cutoff was set manually to 0.00001 for individual CpG probes to determine DMRs.
The most significant CpG site in each DMR was selected and annotated with gene elements (TSS1500, TSS200, 5'UTR, 1st exon, gene body, 3'UTR and intergenic), CpG-island elements (N-shelf, N-shore, CpG-island, S-shelf, S-shore and open sea), and chromatin states based on UCSC genome browser data. Enrichment of the DMRs in a particular genomic region was measured using the Fisher's exact test. DNA motif enrichment in the DMRs was performed using HOMER [41].
The DMRs were mapped to their nearest gene, and pathway analysis was performed using IPA software.

Gene Expression Analyses

Analyses was done using GeneSpring software (Agilent Technologies) to preprocess the gene expression array data according to protocols used before by our group [42-44]. For each expression feature on the array, the log ratio of Cy3 and Cy5 was calculated. The raw log ratios were normalized using baseline transformation and Lowess normalization. The single channel intensities, Cy3 and Cy5, were quantile normalized before analyses. To find differentially expressed array features, a linear model with gender as covariate was build using the limma R package [39].

DNA methylation and gene expression integration

The Spearman correlation was calculated between the methylation value of the most significant Illumina probe in the DMR and the expression levels of the Agilent probes that mapped to their corresponding nearest gene. CpG probe – gene pairs with a correlation p-value below 0.05 were called significant.

CALD1 methylation/expression association with cortisol levels

Linear models controlling for gender were used to calculate the association between CALD1 DNA methylation or gene expression values and cortisol levels and cortisol changes after TSST.
Spearman and Pearson correlations between methylation and gene expression. T-tests were used to examine group differences in methylation levels. Linear regression controlling for gender examined the association between baseline CALD1 expression and cortisol levels. Due to the significant limitations of sample size we did not stratify analyses by gender.

Results

Sample characteristics

The samples differed significantly between groups as expected with greater percent of the child's life spent in institutional care through 12 years of age in the CAUG compared to the FCG. IQ was significantly different between all groups. There were no significant differences in the age at which the assessment occurred between the three groups.

Table 1: Sample characteristics

	CAUG (n=9) Mean (sd)	**FCG (n=7)**	**NIG (N=6)**
Sex	5 males/4 female	3 males/4 females	3 males/3 female
Age	12.02 (.02)	12.16 (.36)	12.22 (.38)
IQ	70.8 (6.9)	80 (7.4)	100 (10.9)
Percent Inst	55.5% (18.5-99.5%), sd 31.9	18.6% (8.9%-52.3%) sd 15.2%	0

Institutional care is associated with DNA methylation changes which are partially remediated by foster care

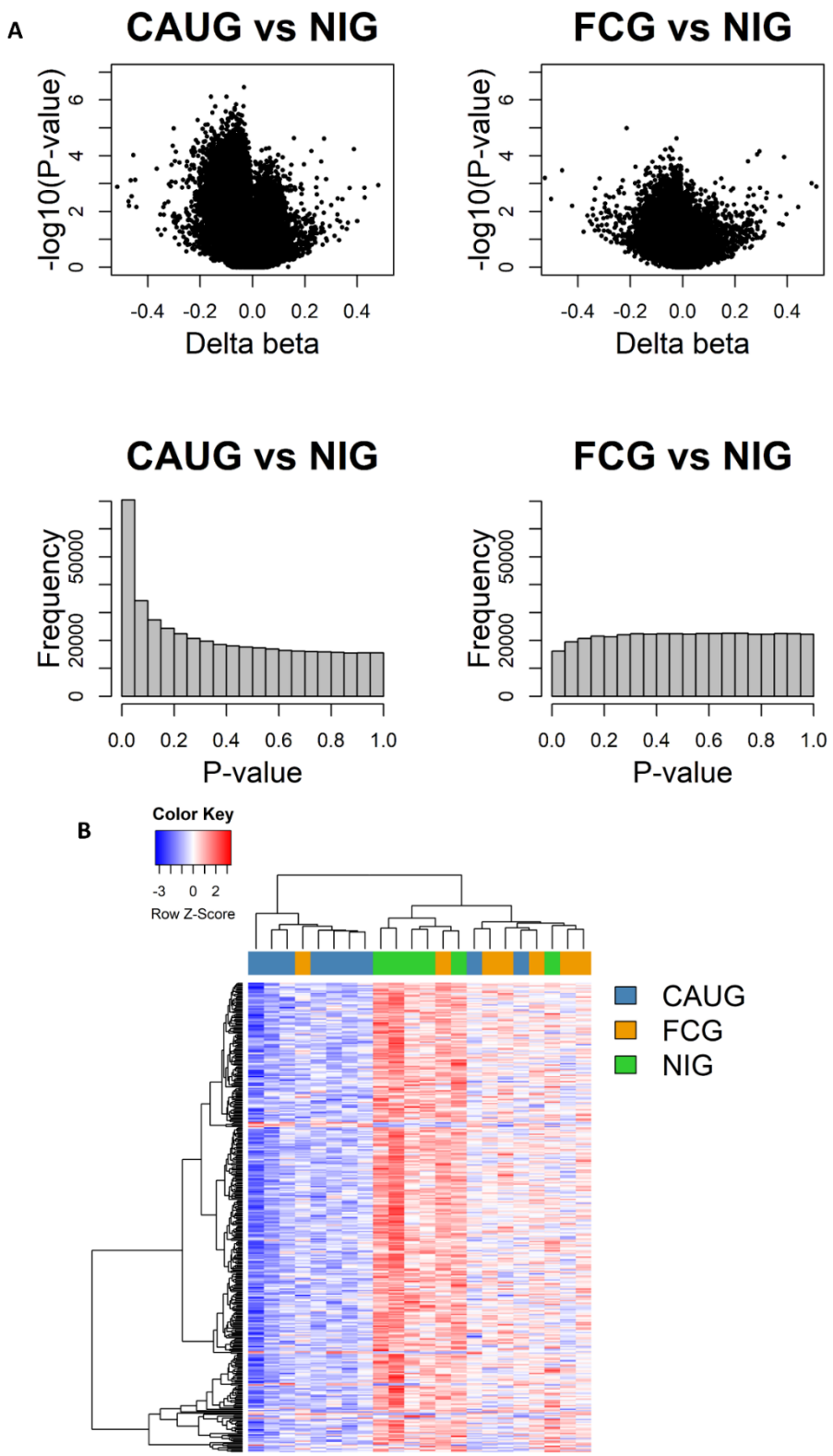
To test whether institutional care may have an epigenetic impact, we measured DNA methylation at >450,000 CpG sites in 9 institutionalized children (CAUG), 7 children in foster care (FCG) and 6 never-institutionalized children (NIG). When compared to NIG, DNA methylation in CAUG and FCG was not found to be significantly different in any of the CpG sites at a false discovery rate (FDR) of 5%. However, upon comparing the distribution of p-values, the effect size was clearly stronger in CAUG compared to FCG as can be seen by the higher density of small p-values in CAUG (**Figure 1A**).

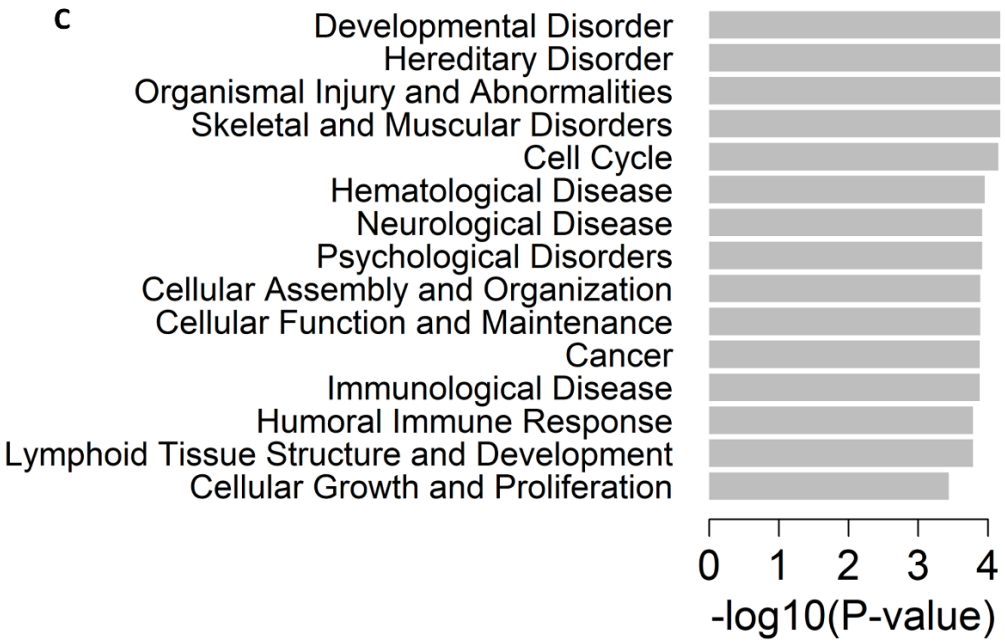
To support these results and simultaneously increase the power to detect significant differences, we focused on DMRs instead of single CpG sites. DMRcate was used to detect DMRs where CpG sites with a p-value lower than 0.00001 were defined as significant. In this way, 367 regions were found to be differentially methylated between CAUG and NIG, while only 6 and 9 DMRs could be found when comparing FCG vs NIG and FCG vs CAUG, respectively (**Supplementary Table 1-3**). Two regions, in the genes *H2AFJ* and *HLA-DRB5*, were both differentially methylated in CAUG and FCG compared to NIG. A heatmap of the most significant CpG sites of the 367 CAUG-DMRs showed a strong hypomethylated profile in CAUG compared to FCG and NIG, where 355 regions were hypomethylated and only 12 regions were hypermethylated (**Figure 1B**).

Furthermore, the epigenetic profiles of children clustered essentially based on their institutional care regimen (FCG, CAUG, NIG). Interestingly, children in FCG had mainly an intermediate methylation profile, with methylation levels in between CAUG and NIG methylation levels, which suggests partial epigenetic remediation of CAUG by FCG.

We next determined whether these DMRs are located at specific genomic regions with respect to gene, CGIs and chromatin states (**Supplementary Table 4**). Compared to all the Illumina CpG probes on the array, the CAUG-DMRs were more likely to be located in gene bodies, with few DMRs near transcription start sites (TSS200, 5'UTR and 1st exon). In addition, the DMRs were found to be depleted in CpG-islands, while enriched in CpG-poor regions outside these islands. We also looked for enrichment in chromatin states which are based on histone modifications resulting in the categorization of genomic regions in 15 cell type-dependent chromatin states, including promoters, enhancers and heterochromatin. We found a strong depletion in promoter states, while CAUG-DMRs were enriched in the transcriptional elongation, weak transcription and heterochromatin states.

We subsequently mapped the DMRs to their nearest gene, and performed IPA pathway analysis. Genes were mainly involved in *developmental disorders*, but also of interest, in *organismal injury and abnormalities*, *neurological diseases* and *psychological disorders* (**Figure 1C**). Of particular interest, we could detect a set of 30 genes which are involved in nervous system and neurons development and function, and are also associated with cognition, behavior and learning (**Figure 1D**). A gene of interest is *catechol-O-methyltransferase* (*COMT*) which is known to play crucial roles in brain function and stress response, and has been linked with different psychological disorders [45, 46]. Furthermore, in a previous study in the same study population, the COMT 158met allele showed protection against depressive symptoms only in the CAU group, which was not seen in the FC group [47].


A
CAUG vs NIG
FCG vs NIG
-log10(P-value)
Delta beta
Frequency
P-value
B
Color Key
-3 0 2
Row Z-Score
CAUG
FCG
NIG


C
Developmental Disorder
Hereditary Disorder
Organismal Injury and Abnormalities
Skeletal and Muscular Disorders
Cell Cycle
Hematological Disease
Neurological Disease
Psychological Disorders
Cellular Assembly and Organization
Cellular Function and Maintenance
Cancer
Immunological Disease
Humoral Immune Response
Lymphoid Tissue Structure and Development
Cellular Growth and Proliferation
0 1 2 3 4
-log10(P-value)

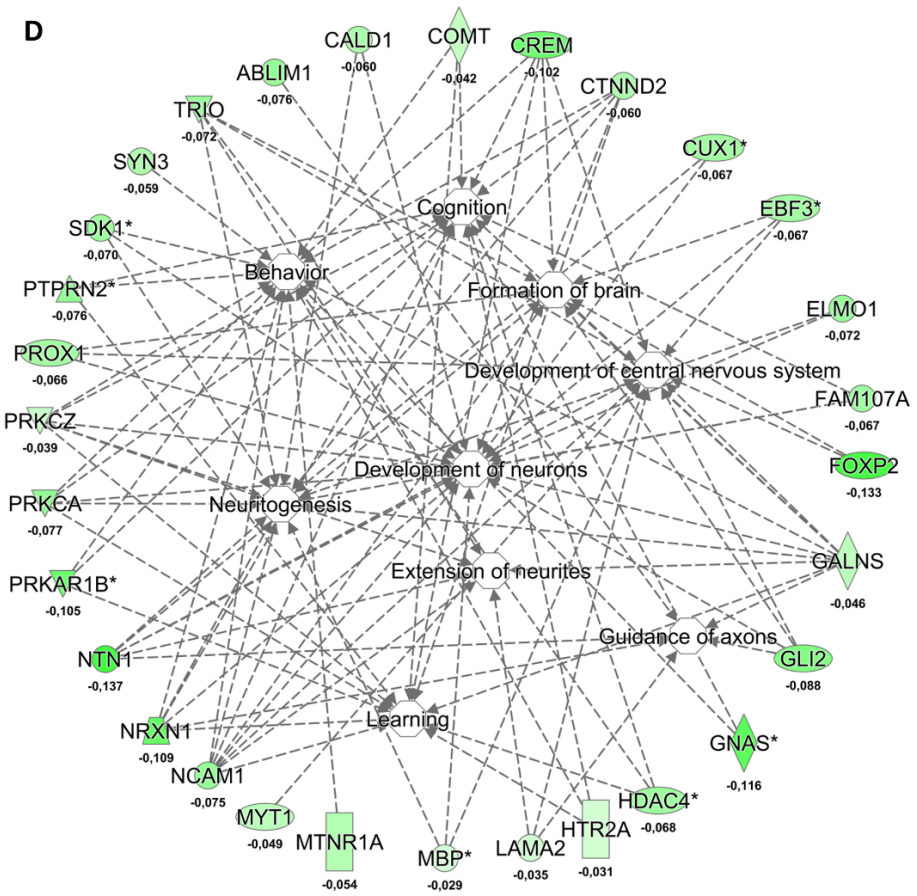

D
CALD1 -0,060
COMT -0,042
CREM -0,102
ABLIM1 -0,076
CTNND2 -0,060
TRIO -0,072
CUX1* -0,067
SYN3 -0,059
Cognition
EBF3* -0,067
SDK1* -0,070
Behavior
Formation of brain
PTPRN2* -0,076
ELMO1 -0,072
PROX1 -0,066
Development of central nervous system
FAM107A -0,067
PRKCZ -0,039
Development of neurons
FOXP2 -0,133
PRKCA -0,077
Neuritogenesis
GALNS -0,046
Extension of neurites
PRKAR1B* -0,105
Guidance of axons
NTN1 -0,137
GLI2 -0,088
Learning
NRXN1 -0,109
GNAS* -0,116
NCAM1 -0,075
HDAC4* -0,068
MYT1 -0,049
HTR2A -0,031
MTNR1A -0,054
MBP* -0,029
LAMA2 -0,035

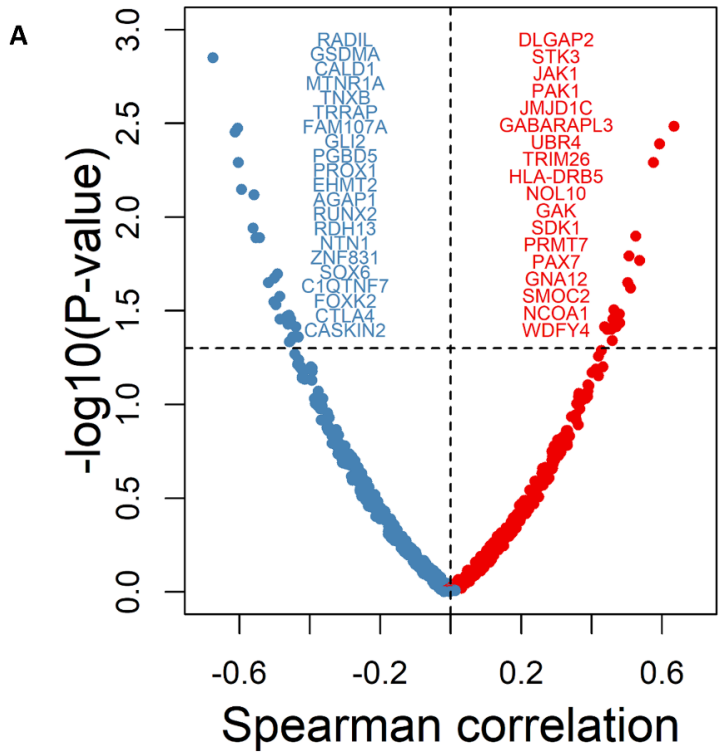
Figure 1: DNA methylation changes associated with institutional care. A) Top row shows volcanoplots of the statistical results comparing CAUG vs NIG and FCG vs NIG. Bottom row shows the p-value distribution in both contrasts. **B)** Heatmap of the methylation levels obtained from the most significant CpG probes in each CAUG-DMR. **C)** Top significant enriched IPA diseases & bio functions. **D)** Genes harboring CAUG-DMRs which are linked with nervous system and neurons development and function, memory, behavior and cognition.

DNA Methylation changes correlated with gene expression are located in genes regulated by nuclear receptors

Because it can be expected that DNA methylation changes modulate gene expression, we compared gene expression profiles before and after a TSST in relation to DNA methylation DMRs. Gene expression changes following TSST, measured as log ratios, were not significantly different across the sample groups (CAUG vs NIG, FCG vs NIG and FCG vs CAUG), although statistical power maybe limited for the small sample size cohorts. Next, we tested whether DNA methylation levels of the CAUG-DMRs were correlated with gene expression by calculation of the spearman correlation between DNA methylation of the most significant CpG probe of each DMR with gene expression log ratios of their corresponding nearest gene. In this way, we could detect 45 CpG probe – gene pairs with a significant correlation (unadjusted p-value < 0.05) between DNA methylation and a change in gene expression after the TSST (log ratio), of which 25 pairs (21 unique genes) were negatively correlated and 20 pairs (18 unique genes) positively (**Figure 2A** and **Supplementary Table 5**). Using IPA, a highly connected network was identified using the correlated genes (**Figure 2B**). Interestingly, different nuclear receptors, including the GR (*NR3C1*), androgen receptor (*AR*), and progesterone receptor (*PGR*) were found to be central network-hubs in the regulatory network.

CALD1 expression and DNA methylation are correlated with cortisol levels

One of the genes with the highest correlation between methylation and gene expression was *CALD1*, a gene with established links to glucocorticoid stress responsiveness, neuronal migration and also the impact of glucocorticoid on neuronal morphology, dendritic spines and migration [48-52]. Given that methylation is expected to influence the gene expression, and particularly the change in gene expression during the experiment of cortisol responsive genes, we tested the association between the fold change in RNA expression with baseline methylation levels of the CpG sites in the DMR. Baseline methylation at the CpG probes cg26816748 and cg15709214 was associated with a gene expression change in three *CALD*1 transcripts, whereas increased methylation was correlated with decreased *CALD1* expression (**Figure 3A**).


A
RADIL
GSDMA
CALD1
MTNR1A
TNXB
TRRAP
FAM107A
GLI2
PGBD5
PROX1
EHMT2
AGAP1
RUNX2
RDH13
NTN1
ZNF831
SOX6
C1QTNF7
FOXK2
CTLA4
CASKIN2
DLGAP2
STK3
JAK1
PAK1
JMJD1C
GABARAPL3
UBR4
TRIM26
HLA-DRB5
NOL10
GAK
SDK1
PRMT7
PAX7
GNA12
SMOC2
NCOA1
WDFY4
-log10(P-value)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
-0.6
-0.2
0.2
0.6
Spearman correlation

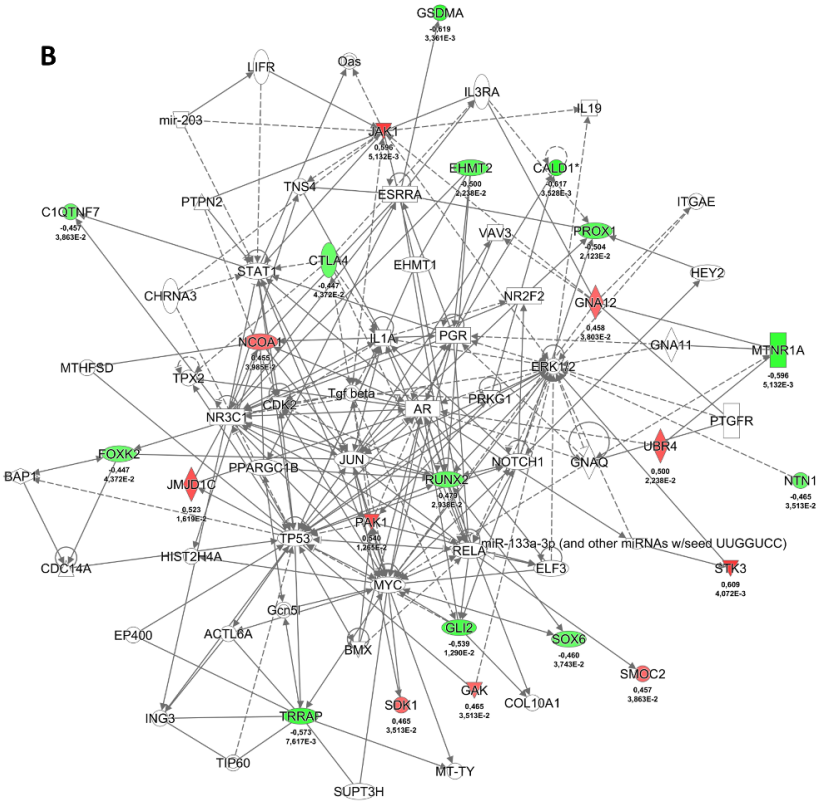
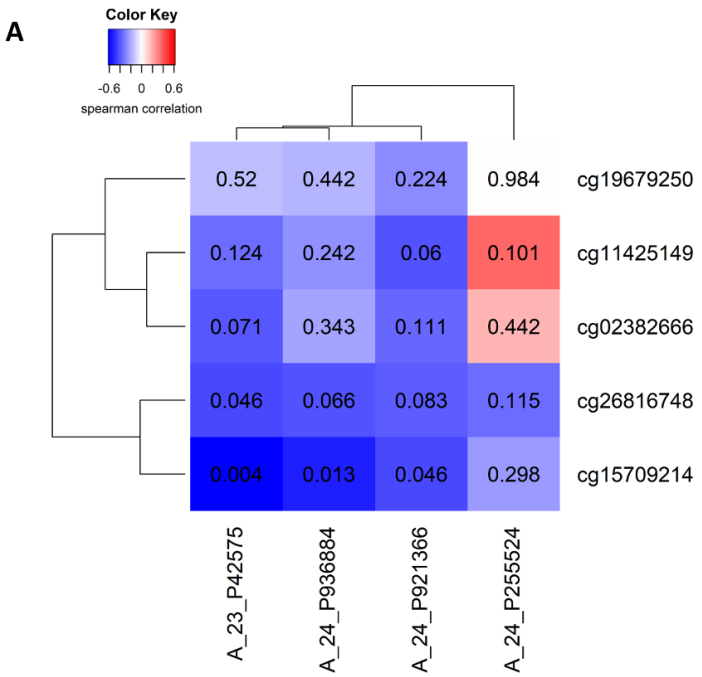
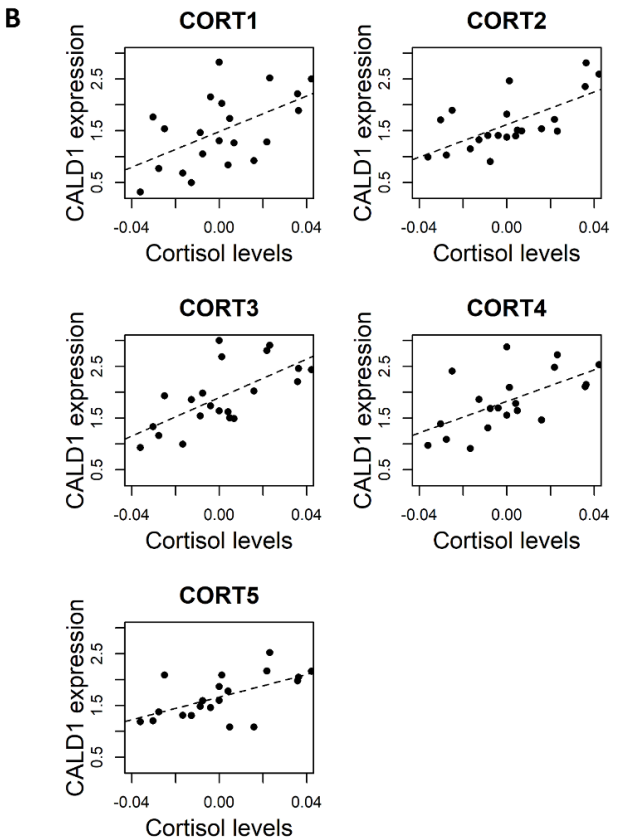

B
GSDMA
LIFR
Oas
IL3RA
mir-203
IL19
JAK1
EHMT2
CALD1
TNS4
ESRRA
PTPN2
ITGAE
C1QTNF7
VAV3
PROX1
CTLA4
EHMT1
HEY2
STAT
CHRNA3
NR2F2
GNA12
NCOA1
IL1A
PGR
GNA11
MTNR1A
MTHFSD
ERK1/2
TPX2
Tgf beta
CDK2
AR
PRKG1
NR3C
PTGFR
FOXK2
JUN
UBR4
PPARGC1B
NOTCH1
GNAQ
BAP1
JMJD1C
RUNX2
NTN1
PAK1
TP53
miR-133a-3p (and other miRNAs w/seed UUGGUCC)
CDC14A
HIST2H4A
RELA
ELF3
STK3
MYC
Gcn5l
EP400
ACTL6A
GLI2
BMX
SOX6
SMOC2
GAK
COL10A1
SOX1
ING3
TRRAP
TIP60
MT-TY
SUPT3H

Figure 2: DNA methylation changes correlated with gene expression. A) Volcano plot showing the results of the Spearman correlation between DNA methylation of the most significant CpG probe in each DMR and their corresponding gene expression logratio values. The significant correlated genes ($p < 0.05$) are represented in blue (negative correlation) and red (positive correlation). **B)** IPA network of the significant correlated genes. Negatively and positively correlated genes are colored in green and red, respectively.

We next tested whether *CALD1* methylation and/or expression was correlated with baseline cortisol levels or changes in cortisol levels after different stress paradigms. We found that baseline *CALD1* expression (A_24_ P921366) was significantly correlated with all cortisol time points (CORT1: p=0.013, beta=15.4; CORT2: p=0.0016, beta=15.3; CORT3: p=0.00052, beta=17; CORT4: p=0.0031, beta=13.3; CORT5: p=0.0098, beta=10.1) (**Figure 3B**), while fold change in *CALD1* expression was not found to be associated with change in cortisol or cortisol levels. Both CALD1 cg15709214 (p=0.03, beta=-0.53) and cg26816748 (p=0.012, beta=-0.54) DNA methylation was associated with baseline cortisol levels (CORT1) only in boys. In addition, the change in cortisol levels between CORT1 and CORT2 was associated with cg2681748 DNA methylation (p=0.036, beta=0.247).

Because *CALD1* expression was previously found to be responsive to glucocorticoids [49, 53, 54], we looked for *NR3C1* DNA motifs (glucocorticoid response elements, GRE) in the CALD1-DMR. Interestingly, a GRE-like motif could be detected just next to CpG probe cg15709214 (**Figure 3C**). Only the 3' cytosine in the GRE was replaced by an adenosine in the *CALD1* sequence.


A
Color Key
-0.6 0 0.6
spearman correlation
0.52 0.442 0.224 0.984 cg19679250
0.124 0.242 0.06 0.101 cg11425149
0.071 0.343 0.111 0.442 cg02382666
0.046 0.066 0.083 0.115 cg26816748
0.004 0.013 0.046 0.298 cg15709214
A_23_P42575
A_24_P936884
A_24_P921366
A_24_P255524


B
CORT1
CORT2
CORT3
CORT4
CORT5
CALD1 expression
Cortisol levels
0.5
1.5
2.5
-0.04
0.00
0.04

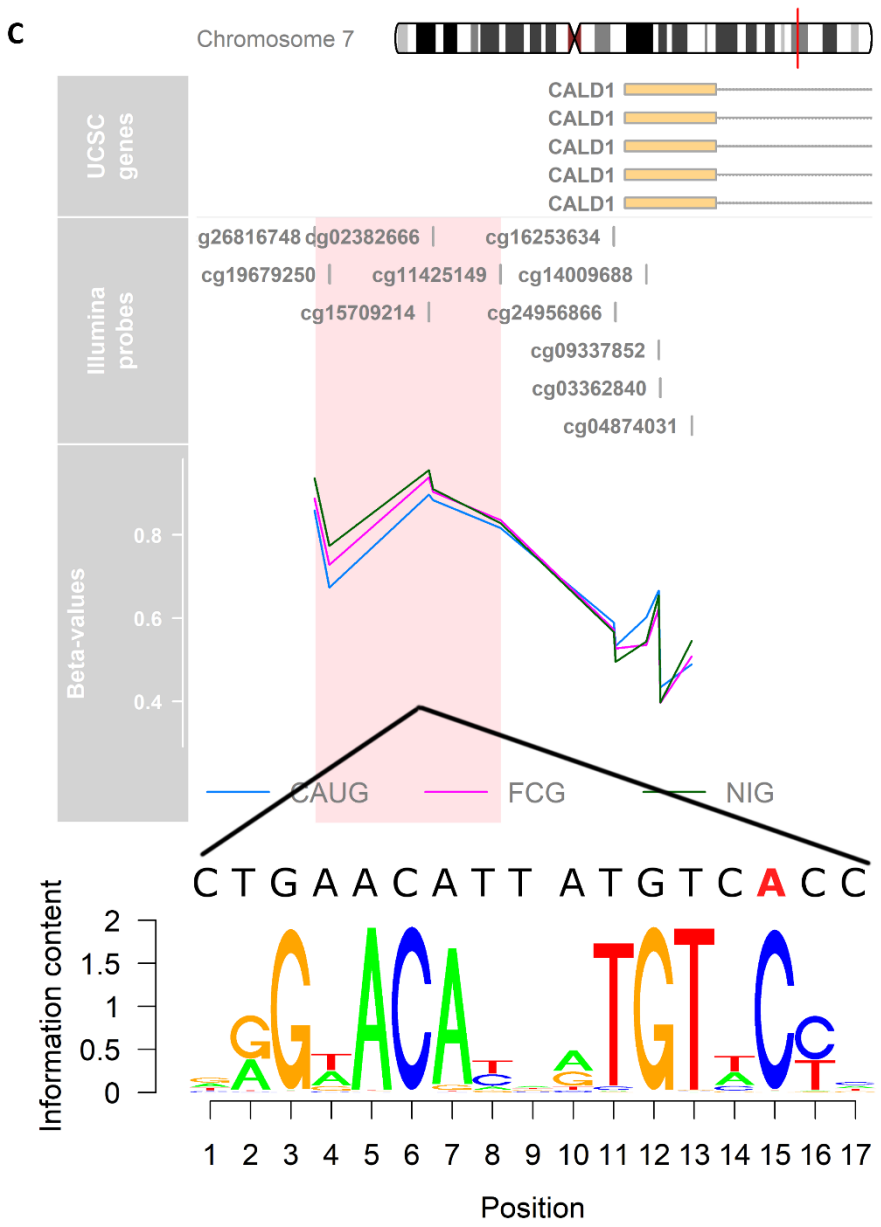


Figure 3: CALD1 methylation/expression correlated with cortisol levels. A) Spearman correlation matrix between CALD1 DMR CpG probes methylation (rows) and CALD1 Agilent probes logratio expression (columns) values. Negative and positive correlations are colored in blue and red, respectively. The pairwise Spearman correlation p-values is given in the heatmap cells. **B)** Scatterplots showing the results of the linear regression between baseline cortisol levels (x-axis) at each timepoint during the social stressor paradigm (CORT1-5) and baseline CALD1 expression (y-axis). **C)** CALD1 promoter region of which the DMR is highlighted in red. The position of the CALD1 Illumina CpG probes are given, and the mean methylation values in each sample group are shown as a line plot. The position and sequence of a GRE-like motif in the DMR is displayed.

Discussion

Our results represent the first study, to our knowledge, to demonstrate, within saliva the association between methylation, gene expression and cortisol. Our stepwise work flow leveraged cross platform analysis of whole genome DNA methylation array (Illumina) and validation in Agilent gene expression array. Uniquely DNA, RNA and cortisol were all obtained from saliva suggesting that the epigenetic regulation of gene expression and their links to a downstream physiologic marker, cortisol, are trackable within saliva offering substantial potential for future studies seeking to define epigenetic mechanistic pathways, particularly those involving vulnerable subjects. We purposefully selected saliva collected during a social evaluative threat paradigm to provide the ability to test the links between three separate salivary analytes. We found a strong hypomethylated profile in children raised in institutions compared to never-institutionalized children. Interestingly, children placed in foster care showed an intermediate DNA methylation pattern, suggesting that foster care may partially remediate these effects. Of particular interest, we could detect a set of 30 DMR associated genes which are involved in nervous system and neuronal development and function, related to cognition, behavior and learning and psychological disorders.

Furthermore, we could correlate a part of these DMRs with corresponding gene expression changes after a TSST paradigm. Of interest, a lot of these genes are under regulation of nuclear receptors including the GR (NR3C1) which plays a central role in stress response. In addition, multiple animal and human studies found an epigenetic association between *NR3C1* regulation and early life stress [55-57]. Further leveraging the randomized control design of the BEIP study and known group differences in salivary reactivity we were able to identify cortisol specific epigenetic control of caldesmon (CALD1) gene expression, with a large body of literature suggesting its mechanistic role in neurodevelopmental alterations as a function of glucocorticoid levels [49, 52].

Upon cross-comparing our data with previous published genome-wide studies on DNA methylation associations with institutional care [12, 14, 15], we found 10 CpG sites which were also differentially methylated in the study of Naumova and coworkers, including *CALD1* (cg02382666) [12] (**Supplementary Table 6**). Also the more recent study of Naumova et al [16] (GSE118940) demonstrates that the most significant sites changed in the same direction as the CAUG-DMRs (**Supplementary Figure 2C and Supplementary Table 9**). Next, we also compared our data with larger sample sized genome-wide studies examining the association between childhood trauma and DNA methylation. A first study studied genome-wide blood Illumina 450k DNA methylation of 85 healthy individuals containing childhood trauma questionnaire (CTQ) scores

(GSE77445) [58]. We performed correlation tests between CTQ scores and DNA methylation of the CAUG-DMRs, and found that the most significant sites changed methylation in the same direction as the CAUG-DMRs (i.e. hypomethylated CAUG-DMRs were predominantly negatively correlated with CTQ) (**Supplementary Figure 2A and Supplementary Table 7**). Similar results could be found in another blood-based genome-wide DNA methylation study of 422 African Americans (GSE72680) (**Supplementary Figure 2B and Supplementary Table 8**). These results further support the validity of our data, despite the low sample size.

CALD1 represents a uniquely important target gene for early life stress response for several reasons. First, in neuronal cultures treatment with high doses GC stress hormone dexamethasone (DEX) results in dose dependent upregulation of *CALD1* expression and subsequently CaD protein. With specific upregulation in the cytosol and the tips of the neurite growth ending of the multipolar processes specifically in the ventricular zone, intermediate zone and cortical plate of the cortex resulting in CaD dependent changes to cell shape and migration [50, 52]. The carboxy terminal domain of CaD negatively regulates interactions between actin and myosin and the change in neuronal precursor cells (NPCs) is mediated by the down regulation of myosin IIa. CaD overexpression results in similar retardation of radial migration as DEX treatment suggesting that GC dependent changes in CaD expression directly impacts radial migration of NPC.

In another study, chronic glucocorticoid stress was found to decrease CaD expression levels in hippocampal neurons, which was found to decrease dendritic spine formation. Combining behavioral analyses with *in vivo* synaptic imaging, it was found that stressful experiences lead to progressive, clustered loss of dendritic neuronal spines and decreased activity of parvalbumin-expressing inhibitory interneurons, which leads do deterioration of sensory functions and text discrimination tasks [48]. Furthermore, glucocorticoids -the end products of HPA-axis- mediate the effects of stress to cause attrition of plasticity in brain regions such as the hippocampus, including simplification of dendrites and shrinkage of dendritic spines. Since CaD expression enlarges the spine-head size by stabilizing F-actin dynamics, CaD expression is a critical target in the GC-induced detrimental effects on dendritic spine development upon chronic stress. Activation of the hypothalamus–pituitary–adrenal (HPA) axis plays a vital role in promoting adaptation during acute stress, but adverse effects of chronic stress may result from overactivity of this system. Remarkably, chronic stress-induced alterations of dendritic spine subtypes triggered functional decrements in an hypothalamus–pituitary–adrenal-inhibitory prefrontal circuit which may dampen its ability to impart inhibitory control over the HPA axis after chronic stress exposure [59].

At the molecular level, conflicting data have been reported with respect to GC dependent CALD1 regulation. In one study in lung cancer cells, *in vitro* and *in vivo* DNA binding assays indicate that GR transactivates the *CALD1* gene by binding directly to the GRE sequence in the promoter [54]. However, in another study in hippocampal neurons, CALD1 promoter reportergene studies show GC dependent repression of the CALD1 promoter which depends on the SRF binding motif, independently of the GRE promoter motifs [49]. Furthermore, in our study, we identified a new GRE-like binding motif near the CALD1 DMR cg15709214 which is associated with transcriptional modulation of 3 CaD transcript variants and cortisol levels.

Together, these findings indicate an important mechanistic role for CALD1 in the lasting neurodevelopmental impacts of early life stress and suggest that changes in *CALD1* methylation and resulting alterations in gene expression and CaD isoform protein production may contribute to the adverse lasting consequences associated with institutional care stress across social, emotional, behavioral, physiological, neurobiological and health outcomes.

Despite the unique nature of these findings there are significant limitations. The first notable limitation is the small sample size, given the use of two different arrays and the significant potential for false positive findings our results should be interpreted as pilot data and proof of concept. Recognizing the unique nature of salivary gene expression, the cost of expression and methylation arrays, and the need to demonstrate proof of concept analyses was done only on a subset of samples. That being noted, it remains significant that the gene identified through this workflow has established links to cortisol responsivity, early and prenatal stress and neurodevelopment indicating that spurious associations are unlikely. Second, methylation significantly predicted fold change in gene expression and not baseline levels. Although methylation is typically thought to be associated with gene expression our findings may be explained by the use of the TSST, a paradigm in which we expected to see changes in cortisol levels during the TSST and importantly also detect group differences in the reactivity patterns of cortisol. As such our workflow design was more likely to identify genes that changed over the course of the TSST rather than those predicting baseline results. An alternative more mechanistic explanation is that if methylation influenced the accessibility of the GRE in *CALD1* our finding would be in line with GC induced expression of *CALD1* and suggest that at baseline *CALD1* expression does not have a significant link to cortisol levels. To support this statement, we found a GRE-like motif inside the *CALD1* DMR, which was different from the two GRE sequences closer towards the transcription start site of *CALD1*. Further molecular experiments should reveal whether GR is able to bind at this GRE and subsequently leads to changes in *CALD1* gene expression which controls neuronal migration and dendritic spine formation.

The exact mechanism by which signals from the early caregiving environment impinge on the developing brain to shape underlying neural pathways and circuits and what role epigenetic processes have in directing these trajectories, both positive and negative, is not yet fully elucidated [14]. An increasing body of research supports the utility of DNA methylation in peripheral tissues as biomarkers of exposure and potential predictive factors linked to later neuropsychological outcomes [60]. In rare case central tissue is available to draw causative relations but, for the most part, epigenetic studies in humans will continue to be limited to peripheral samples particularly in longitudinal prospective study designs. Given that limitation, maximizing the information related to causal pathways that can be determined from easily obtainable, and easily replenished, biospecimens like saliva represents an important next step for human studies seeking to define the mechanisms linking experience and phenotypes at any age. Our results offer an exciting glimpse into the potential utility of salivary diagnostics to define the underlying mechanisms through which early life adversity is biologically embedded and perhaps even more importantly a unique opportunity in which to examine the impact of interventions on biological processes without invasive protocols and with methodology that can be utilized in remote locations and processed at later time points.

Acknowledgements

PDB is recipient of a Bill & Melinda Gates Foundation Grand Challenges Exploration Grant (OPP119403) in the field of saliva biomarker discovery, enabling collaboration with SD. WVB and KDK are supported by FWO grants G079614N and G059713N and BOF NOI/DOCPRO grants (UA).

References

1. Nelson, C.A., 3rd, et al., *Cognitive recovery in socially deprived young children: the Bucharest Early Intervention Project.* Science, 2007. **318**(5858): p. 1937-40.
2. Fagiolini, M., C.L. Jensen, and F.A. Champagne, *Epigenetic influences on brain development and plasticity.* Curr Opin Neurobiol, 2009. **19**(2): p. 207-12.
3. Heim, C. and E.B. Binder, *Current research trends in early life stress and depression: review of human studies on sensitive periods, gene-environment interactions, and epigenetics.* Exp Neurol, 2012. **233**(1): p. 102-11.
4. Gudsnuk, K. and F.A. Champagne, *Epigenetic influence of stress and the social environment.* ILAR J, 2012. **53**(3-4): p. 279-88.
5. Boyce, W.T. and M.S. Kobor, *Development and the epigenome: the 'synapse' of gene-environment interplay.* Dev Sci, 2015. **18**(1): p. 1-23.
6. Drury, S.S., M.M. Sanchez, and A. Gonzalez, *When mothering goes awry: Challenges and opportunities for utilizing evidence across rodent, nonhuman primate and human studies to better define the biological consequences of negative early caregiving.* Horm Behav, 2016. **77**: p. 182-92.
7. Roth, T.L., et al., *Bdnf DNA methylation modifications in the hippocampus and amygdala of male and female rats exposed to different caregiving environments outside the homecage.* Dev Psychobiol, 2014. **56**(8): p. 1755-63.
8. Yang, B.Z., et al., *Child abuse and epigenetic mechanisms of disease risk.* Am J Prev Med, 2013. **44**(2): p. 101-7.
9. Provencal, N., et al., *The signature of maternal rearing in the methylome in rhesus macaque prefrontal cortex and T cells.* J Neurosci, 2012. **32**(44): p. 15626-42.
10. Zhang, T.Y. and M.J. Meaney, *Epigenetics and the environmental regulation of the genome and its function.* Annu Rev Psychol, 2010. **61**: p. 439-66, C1-3.
11. Nieratschker, V., et al., *MORC1 exhibits cross-species differential methylation in association with early life stress as well as genome-wide association with MDD.* Transl Psychiatry, 2014. **4**: p. e429.
12. Naumova, O.Y., et al., *Differential patterns of whole-genome DNA methylation in institutionalized children and children raised by their biological parents.* Dev Psychopathol, 2012. **24**(1): p. 143-55.
13. Non, A.L., et al., *DNA methylation at stress-related genes is associated with exposure to early life institutionalization.* Am J Phys Anthropol, 2016. **161**(1): p. 84-93.
14. Kumsta, R., et al., *Severe psychosocial deprivation in early childhood is associated with increased DNA methylation across a region spanning the transcription start site of CYP2E1.* Transl Psychiatry, 2016. **6**(6): p. e830.
15. Esposito, E.A., et al., *Differential DNA methylation in peripheral blood mononuclear cells in adolescents exposed to significant early but not later childhood adversity.* Dev Psychopathol, 2016. **28**(4pt2): p. 1385-1399.
16. Naumova, O.Y., et al., *Effects of early social deprivation on epigenetic statuses and adaptive behavior of young children: A study based on a cohort of institutionalized infants and toddlers.* PLoS One, 2019. **14**(3): p. e0214285.
17. Gurnot, C., et al., *Prenatal antidepressant exposure associated with CYP2E1 DNA methylation change in neonates.* Epigenetics, 2015. **10**(5): p. 361-72.
18. Zeanah, C.H., et al., *Designing research to study the effects of institutionalization on brain and behavioral development: the Bucharest Early Intervention Project.* Dev Psychopathol, 2003. **15**(4): p. 885-907.
19. Fox, N.A., et al., *The effects of severe psychosocial deprivation and foster care intervention on cognitive development at 8 years of age: findings from the Bucharest Early Intervention Project.* J Child Psychol Psychiatry, 2011. **52**(9): p. 919-28.
20. McLaughlin, K.A., et al., *Causal effects of the early caregiving environment on development of stress response systems in children.* Proc Natl Acad Sci U S A, 2015. **112**(18): p. 5637-42.

21. Wren, M.E., E.A. Shirtcliff, and S.S. Drury, *Not all biofluids are created equal: chewing over salivary diagnostics and the epigenome.* Clin Ther, 2015. **37**(3): p. 529-39.
22. Zimmermann, B.G., N.J. Park, and D.T. Wong, *Genomic targets in saliva.* Ann N Y Acad Sci, 2007. **1098**: p. 184-91.
23. Klengel, T., et al., *Allele-specific FKBP5 DNA demethylation mediates gene-childhood trauma interactions.* Nat Neurosci, 2013. **16**(1): p. 33-41.
24. Boks, M.P., et al., *SKA2 Methylation is Involved in Cortisol Stress Reactivity and Predicts the Development of Post-Traumatic Stress Disorder (PTSD) After Military Deployment.* Neuropsychopharmacology, 2016. **41**(5): p. 1350-6.
25. Mulder, R.H., et al., *Methylation matters: FK506 binding protein 51 (FKBP5) methylation moderates the associations of FKBP5 genotype and resistant attachment with stress regulation.* Dev Psychopathol, 2017. **29**(2): p. 491-503.
26. Labonte, B., et al., *Epigenetic modulation of glucocorticoid receptors in posttraumatic stress disorder.* Transl Psychiatry, 2014. **4**: p. e368.
27. Hicks, S.D., et al., *Salivary miRNA profiles identify children with autism spectrum disorder, correlate with adaptive behavior, and implicate ASD candidate genes involved in neurodevelopment.* BMC Pediatr, 2016. **16**: p. 52.
28. Kirschbaum, C. and D.H. Hellhammer, *Salivary cortisol in psychobiological research: an overview.* Neuropsychobiology, 1989. **22**(3): p. 150-69.
29. Vining, R.F., et al., *Salivary cortisol: a better measure of adrenal cortical function than serum cortisol.* Ann Clin Biochem, 1983. **20 (Pt 6)**: p. 329-35.
30. Zeanah, C.H., et al., *Institutional rearing and psychiatric disorders in Romanian preschool children.* Am J Psychiatry, 2009. **166**(7): p. 777-85.
31. Smyke, A.T., et al., *The caregiving context in institution-reared and family-reared infants and toddlers in Romania.* J Child Psychol Psychiatry, 2007. **48**(2): p. 210-8.
32. Zeanah, C.H., N.A. Fox, and C.A. Nelson, *The Bucharest Early Intervention Project: case study in the ethics of mental health research.* J Nerv Ment Dis, 2012. **200**(3): p. 243-7.
33. Langie, S.A.S., et al., *GLI2 promoter hypermethylation in saliva of children with a respiratory allergy.* Clin Epigenetics, 2018. **10**: p. 50.
34. Assenov, Y., et al., *Comprehensive analysis of DNA methylation data with RnBeads.* Nat Methods, 2014. **11**(11): p. 1138-1140.
35. van Iterson, M., et al., *MethylAid: visual and interactive quality control of large Illumina 450k datasets.* Bioinformatics, 2014. **30**(23): p. 3435-7.
36. Teschendorff, A.E., et al., *A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data.* Bioinformatics, 2013. **29**(2): p. 189-96.
37. Houseman, E.A., et al., *DNA methylation arrays as surrogate measures of cell mixture distribution.* BMC Bioinformatics, 2012. **13**: p. 86.
38. Reinius, L.E., et al., *Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility.* PLoS One, 2012. **7**(7): p. e41361.
39. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and microarray studies.* Nucleic Acids Res, 2015. **43**(7): p. e47.
40. Peters, T.J., et al., *De novo identification of differentially methylated regions in the human genome.* Epigenetics Chromatin, 2015. **8**: p. 6.
41. Heinz, S., et al., *Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities.* Mol Cell, 2010. **38**(4): p. 576-89.
42. De Boever, P., et al., *Perinatal exposure to purity-controlled polychlorinated biphenyl 52, 138, or 180 alters toxicogenomic profiles in peripheral blood of rats after 4 months.* Chem Res Toxicol, 2013. **26**(8): p. 1159-67.
43. De Boever, P., et al., *Characterization of the peripheral blood transcriptome in a repeated measures design using a panel of healthy individuals.* Genomics, 2014. **103**(1): p. 31-9.

44. Wens, B., et al., *Transcriptomics identifies differences between ultrapure non-dioxin-like polychlorinated biphenyls (PCBs) and dioxin-like PCB126 in cultured peripheral blood mononuclear cells.* Toxicology, 2011. **287**(1-3): p. 113-23.
45. Witte, A.V. and A. Floel, *Effects of COMT polymorphisms on brain function and behavior in health and disease.* Brain Res Bull, 2012. **88**(5): p. 418-28.
46. Hosak, L., *Role of the COMT gene Val158Met polymorphism in mental disorders: a review.* Eur Psychiatry, 2007. **22**(5): p. 276-81.
47. Drury, S.S., et al., *Modification of depression by COMT val158met polymorphism in children exposed to early severe psychosocial deprivation.* Child Abuse Negl, 2010. **34**(6): p. 387-95.
48. Chen, C.C., et al., *Selective activation of parvalbumin interneurons prevents stress-induced synapse loss and perceptual defects.* Mol Psychiatry, 2018. **23**(7): p. 1614-1625.
49. Tanokashira, D., et al., *Glucocorticoid suppresses dendritic spine development mediated by down-regulation of caldesmon expression.* J Neurosci, 2012. **32**(42): p. 14583-91.
50. Fukumoto, K., et al., *Detrimental effects of glucocorticoids on neuronal migration during brain development.* Mol Psychiatry, 2009. **14**(12): p. 1119-31.
51. Xiong, F. and L. Zhang, *Role of the hypothalamic-pituitary-adrenal axis in developmental programming of health and disease.* Front Neuroendocrinol, 2013. **34**(1): p. 27-46.
52. Sobue, K. and K. Fukumoto, *Caldesmon, an actin-linked regulatory protein, comes across glucocorticoids.* Cell Adh Migr, 2010. **4**(2): p. 185-9.
53. Anderson, R.M., et al., *Prolonged corticosterone exposure induces dendritic spine remodeling and attrition in the rat medial prefrontal cortex.* J Comp Neurol, 2016. **524**(18): p. 3729-3746.
54. Mayanagi, T., et al., *Glucocorticoid receptor-mediated expression of caldesmon regulates cell migration via the reorganization of the actin cytoskeleton.* J Biol Chem, 2008. **283**(45): p. 31183-96.
55. Weaver, I.C., et al., *Epigenetic programming by maternal behavior.* Nat Neurosci, 2004. **7**(8): p. 847-54.
56. McGowan, P.O., et al., *Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse.* Nat Neurosci, 2009. **12**(3): p. 342-8.
57. Palma-Gudiel, H., et al., *Glucocorticoid receptor gene (NR3C1) methylation processes as mediators of early adversity in stress-related disorders causality: A critical review.* Neurosci Biobehav Rev, 2015. **55**: p. 520-35.
58. Houtepen, L.C., et al., *Genome-wide DNA methylation levels and altered cortisol stress reactivity following childhood trauma in humans.* Nat Commun, 2016. **7**: p. 10967.
59. Radley, J.J., et al., *Chronic stress-induced alterations of dendritic spine subtypes predict functional decrements in an hypothalamo-pituitary-adrenal-inhibitory prefrontal circuit.* J Neurosci, 2013. **33**(36): p. 14379-91.
60. Hannon, E., et al., *Interindividual methylomic variation across blood, cortex, and cerebellum: implications for epigenetic studies of neurological and neuropsychiatric phenotypes.* Epigenetics, 2015. **10**(11): p. 1024-32.

Supplementary information

Supplementary tables 1 to 9 can be found in following dropbox folder: https://www.dropbox.com/sh/d9lq9mipqkoezx3/AAC9RCXgWS7P6DS71ybJqWNJa?dl=0

Supplementary Table 1: Differentially methylated regions comparing CAUG with NIG.

Supplementary Table 2: Differentially methylated regions comparing FCG with NIG.

Supplementary Table 3: Differentially methylated regions comparing FCG with CAUG.

Supplementary Table 4: Enrichment of CAUG-DMRs in different genomic locations.

Supplementary Table 5: Correlation between DNA methylation and gene expression of the CAUG-DMRs.

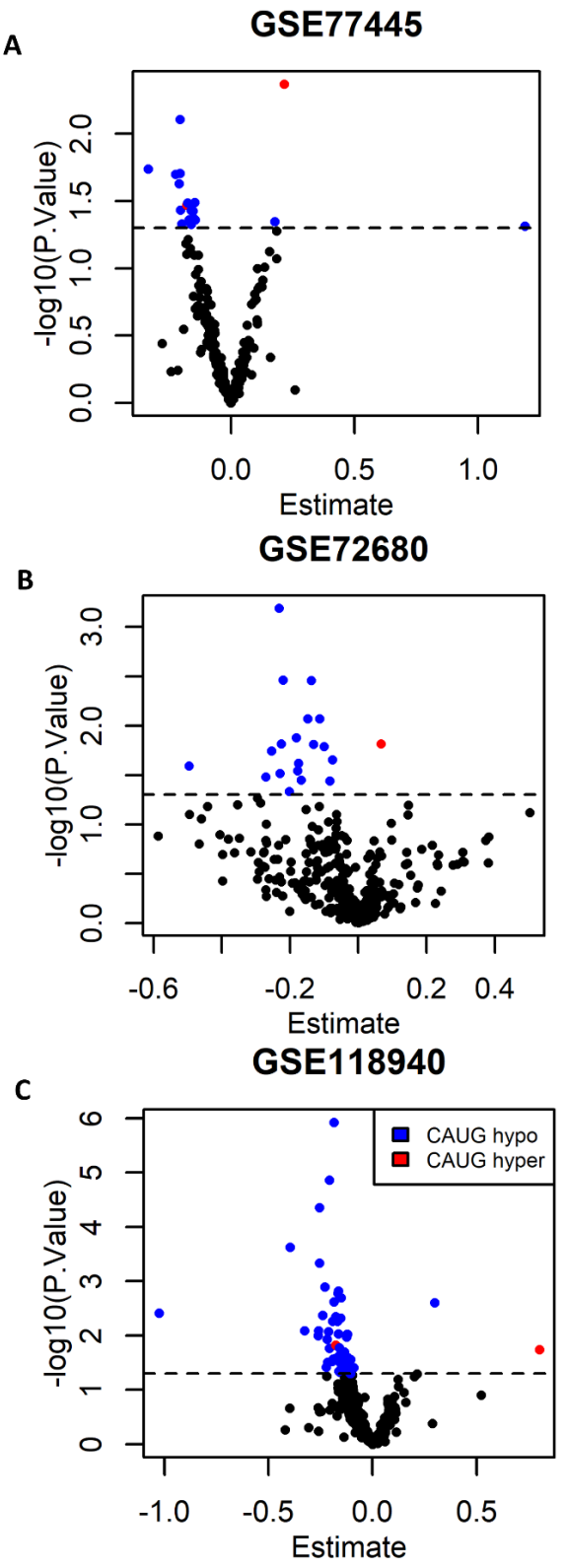
Supplementary Table 6: Overlapping differentially methylated CpG sites between CAUG-DMRs and study of Naumova et al. [12].

Supplementary Table 7: Overlapping differentially methylated CpG sites between CAUG-DMRs and GSE77445 dataset.

Supplementary Table 8: Overlapping differentially methylated CpG sites between CAUG-DMRs and GSE72680 dataset.

Supplementary Table 9: Overlapping differentially methylated CpG sites between CAUG-DMRs and GSE118940 dataset.

Supplementary Figure 1: Timing of cortisol and RNA collections.


A
GSE77445
-log10(P.Value)
0.0
0.5
1.0
1.5
2.0
0.0
0.5
1.0
Estimate
B
GSE72680
-log10(P.Value)
0.0
1.0
2.0
3.0
-0.6
-0.2
0.2
0.4
Estimate
C
GSE118940
CAUG hypo
CAUG hyper
-log10(P.Value)
0
1
2
3
4
5
6
-1.0
-0.5
0.0
0.5
Estimate

Supplementary Figure 2: Replication of CAUG-DMRs in other genome-wide DNA methylation cohorts. Volcanoplots showing the results of the linear regression tests associating DNA methylation with CTQ scores. CpG probes colored in blue and red are significant associated with CTQ, and hypo- and hypermethylated in CAUG, respectively. Results of GEO datasets A) GSE77445, B) GSE72680 and C) GSE118940.

5

Identification of differentially methylated BRCA1 and CRISP2 DNA regions as blood surrogate markers for cardiovascular disease

Ken Declerck[1#], Geoffrey Istas[1,2,3#], Maria Pudenz[4], Katarzyna Szarc vel Szic[1,5], Veronica Lendinez-Tortajada[6], Montserrat Leon-Latre[7,8], Karen Heyninck[9], Guy Haegeman[9], Jose A. Casasnovas[8,10,11], Maria Tellez-Plaza[12], Clarissa Gerhauser[4], Christian Heiss[2], Ana Rodriguez-Mateos[2,3], Wim Vanden Berghe[1,9]

[1] Laboratory of Protein chemistry, Proteomics and Epigenetic Signaling (PPES), Department of Biomedical Sciences, Faculty of Pharmaceutical, Biomedical and Veterinary Sciences, Antwerp University, Antwerp (Wilrijk), Belgium
[2] Division of Cardiology, Pulmonology, and Vascular Medicine, Medical Faculty, Düsseldorf University, Düsseldorf, Germany
[3] Current affiliation:: Division of Diabetes and Nutritional Sciences, Faculty of Life Sciences and Medicine, King's College London
[4] Workgroup Cancer Chemoprevention and Epigenomics, Division of Epigenomics and Cancer Risk Factors, German Cancer Research Center (DKFZ), Heidelberg, Germany
[5] Division of Hematology, Oncology and Stem Cell Transplantation, Center for Translational Cell Research, The University Medical Center Freiburg, Freiburg, Germany
[6] Genomic and Genetic Diagnosis Unit, Institute for Biomedical Research Hospital Clinic de Valencia, Valencia, Spain
[7] Servicio Aragones de Salud, Zaragoza, Spain
[8] IIS de Aragon, Zaragoza, Spain
[9] Laboratory of Eukaryotic Gene Expression and Signal Transduction LEGEST, Department of Biochemistry and Microbiology, Ghent University, Gent, Belgium

[10] Instituto Aragonés de Ciencias de Salud, Zaragoza, Spain
[11] Universidad de Zaragoza, Zaragoza, Spain
[12] Workgroup Cardiometabolic and Renal Risk, Institute for Biomedical Research Hospital Clinic de Valencia, Valencia, Spain
[#] equally contributed

Published in Scientific Reports (https://doi.org/10.1038/s41598-017-03434-0)

Contributions: Ken Declerck analyzed the data, interpreted the results and wrote, together with Geoffrey Istas, the manuscript.

Abstract | Genome-wide Illumina InfiniumMethylation 450K DNA methylation analysis was performed on blood samples from clinical atherosclerosis patients (n=8) and healthy donors (n=8) in the LVAD study (NCT02174133, NCT01799005). Multiple DMRs could be identified in atherosclerosis patients, related to epigenetic control of cell adhesion, chemotaxis, cytoskeletal reorganizations, cell proliferation, cell death, estrogen receptor pathways and phagocytic immune responses. Furthermore, a subset of 34 DMRs related to impaired oxidative stress, DNA repair, and inflammatory pathways could be replicated in an independent cohort study of donor-matched healthy and atherosclerotic human aorta tissue (n=15) and human carotid plaque samples (n=19). Upon integrated network analysis, BRCA1 and CRISP2 DMRs were identified as most central disease-associated DNA methylation biomarkers. Differentially methylated BRCA1 and CRISP2 regions were verified by MassARRAY Epityper and pyrosequencing assays and could be further replicated in blood, aorta tissue and carotid plaque material of atherosclerosis patients. Moreover, methylation changes at BRCA1 and CRISP2 specific CpG sites were consistently associated with subclinical atherosclerosis measures (coronary calcium score and carotid intima media thickness) in an independent sample cohort of middle-aged men with subclinical CVD in the Aragon Workers' Health Study (n=24). Altogether, BRCA1 and CRISP2 DMRs hold promise as novel blood surrogate markers for early risk stratification and CVD prevention.

Introduction

According to the World Health Organization (WHO), CVDs account for the highest mortality numbers with approximately 30% of all deaths worldwide (http://www.who.int/gho/ncd/en/). Atherosclerosis is the major principle underlying CVD. At predisposed areas of the vascular tree, including the branching points of coronary and carotid arteries, localized accumulation of fatty deposits and inflammation reactions contribute to plaque development and progression eventually leading to impaired blood flow resulting in CVD i.e. coronary artery disease and cerebrovascular disease [1].

The development of an atherosclerotic lesion is a slow and silent process making early stage diagnosis difficult [2]. Early detection of individuals in the process of developing atherosclerosis might be essential for cardiovascular prevention. Approximately 60% of individuals categorized as at low risk for CVD based on traditional risk factors prediction equations had subclinical atherosclerosis [3, 4]. Thus, other factors not traditionally included in risk scales are likely to be involved in atherogenesis.

Preclinical evidence supports that aberrant monocyte-macrophage differentiation contributes to vascular wall inflammation in patients at high risk for atherosclerosis [5, 6]. CpG DNA methylation is involved in the epigenetic differentiation and regulation of leukocyte specific gene expression profiles, including the expression of soluble mediators and surface molecules that direct margination, adhesion, and migration of blood leukocytes in vascular tissues [7]. While very little is known about the human leukocyte DNA methylome and its potential causal role in CVD, blood DNA methylation markers may contribute to the diagnosis of atherosclerosis patients. Recent studies have illustrated the feasibility of DNA methylation profiling using peripheral blood to identify CVD specific surrogate biomarkers [8-15]. Candidate-gene approaches identified significant associations between leukocyte DNA methylation and atherosclerosis, whereas the results for the association between global DNA methylation and atherosclerosis were not always consistent [10, 11, 14, 15]. Previous studies however, evaluated differentially methylated sites using samples from individuals with clinical CVD, but did not examine the potential role of DNA methylation regions as a marker of subclinical disease.

Therefore, we characterized genome-wide DNA methylation profiles of blood samples of atherosclerosis patients versus healthy individuals from the LVAD study (Impact of Left Ventricular Assist Devices Implantation on Micro- and Macrovascular Function, NCT02174133) and identified promising atherosclerosis-related epigenetic biomarkers. For the selected regions, we validated these whole blood DNA methylation profiles using publicly available Illumina InfiniumMethylation 450K data from carotid normal and atherosclerotic plaque samples. Additionally, we compared CVD associated DNA methylation changes with aging and/or immune cell epigenotypes. Finally, we explored the role of promising regions as potential predictors of subclinical atherosclerosis in a subsample of 24 individuals that participated in the Aragon Workers Health Study (AWHS). The AWHS is a prospective cohort that aims to characterize the factors associated with metabolic abnormalities and imaging-based subclinical atherosclerosis measures in a middle-aged population free of clinical CVD [3, 16].

Materials and methods

We fist conducted a discovery phase analysis of DNA methylation data from 8 healthy voluntaries and 8 patients with atherosclerosis from the Study “Impact of Left Ventricular Assist Devices Implantation on Micro- and Macrovascular Function” (LVAD study, clinicaltrials.gov: NCT02174133). Subsequently we validated findings from the discovery phase by analyzing DNA methylation data from plaque material related to GEO dataset GSE46401 published by Zaina et al. Finally, as a post-hoc analysis, we

explored the potential role of the identified markers as early detection biomarkers by evaluating the association of DNA methylation and subclinical atherosclerosis endpoints in whole blood DNA from 24 Aragon Workers Study participants free of clinical CVD.

Experimental set-up and sample collection in the discovery stage

In the LVAD study, eight healthy volunteers were recruited based on following inclusion criteria: age (35 – 60 years), BMI (23-27 kg/m²), average physical activity and normal western diet (clinicaltrials.gov: NCT01799005). The exclusion criteria for the healthy volunteers were CVD, diabetes mellitus, acute inflammation and arrhythmia. We additionally selected eight patients with confirmed clinical diagnosis of atherosclerosis (clinicaltrials.gov: NCT02174133). The characteristics of the study population are described in table 1. Whole blood (0.5 ml) was collected from all individuals, following informed consent. The study was conducted according to the guidelines laid down in the Declaration of Helsinki and all procedures involving human subjects were approved by the University of Düsseldorf Research Ethics Committee (ref: 3870 and ref: 4565R).

Table 1: Volunteer characteristics

	Healthy Mean ± SD	Atherosclerosis Mean ± SD	P value
Coronary artery disease	No	Yes	
Arterial hypertension	No	Yes	
Arrhythmia	No	Yes	
Age	49.7 ± 6.6	78.0 ± 8.4	<0.0001
BMI (kg/m2)	26.2 ± 2.3	27.0 ± 3.2	0.61
LDL cholesterol (mg/dl)	147.4 ± 37.2	121.3 ± 14.5	0.13
HDL cholesterol (mg/dl)	52.3 ± 11.2	46.8 ± 10.1	0.41
Total cholesterol (mg/dl)	212.7 ± 38.0	179.0 ± 32.5	0.09
Fasting plasma glucose (mg/dl)	90.0 ± 8.0	86.1 ± 14.3	0.54
HbA1c (%)	5.7 ± 0.3	5.8 ± 0.9	0.78
CRP (mg/dl)	0.6 ± 0.5	0.7 ± 0.5	0.04
Leukocytes (1000/ul)	6.6 ± 2.3	6.4 ± 1.4	0.84
Hb (mg/dl)	12.7 ± 1.7	14.9 ± 0.6	0.007
Creatinine (mg/dl)	1.0 ± 0.2	1.3 ± 0.5	0.31
Triglycerides (mg/dl)	95.9 ± 44.2	159.5 ± 48.0	0.03

Statistical analysis performed by means of the student T test.

Infinium HumanMethylation450 BeadChip Array processing and data analysis for whole blood DNA from atherosclerosis patients and healthy donors

gDNA isolated from 0.5 ml whole blood (EDTA), was isolated with DNeasy Blood & Tissue kit (Qiagen Hilden, Germany) and quantified by Nanodrop™ spectrophotometry. 1000 ng of gDNA was bisulfite converted using the EZ DNA methylation kit (Zymo Research, Orange, CA, USA) according to manufacturer's instructions. Genome-wide DNA methylation was analyzed on Infinium HumanMethylation450 BeadChip platform (Illumina, San Diego, CA, USA) at the DKFZ Genomics and Proteomics Core Facility. 4 µl of bisulfite-converted whole blood DNA (~150 ng) was used for the whole genome amplification (WGA) reaction, enzymatic fragmentation, precipitation and re-suspended in hybridization buffer. Subsequent steps of DNA methylation analysis were carried out according to the standard Infinium HD Assay Methylation Protocol Guide (Part #15019519, Illumina). The BeadChip images were captured using the Illumina iScan. Pre-processing and analysis of the Infinium 450k data was performed using the R package RnBeads [17]. CpG probes containing a SNP at least 3 bp from the 3' query site, having a detection p-value higher than 0.01, having empty values in at least one sample or measuring methylation in a non-CpG context were removed. In total 8,533 CpG probes (1.75%) were filtered. Intra-array normalization was done using the Beta Mixture Quantile Normalization [18]. Methylation values were represented as β-values ranging from 0 to 1. β-values were converted into M-values ($M = \log2(\beta/((1-\beta)))$) before doing the statistical analysis. Limma R package was used to identify DMPs. Raw p-values were corrected for multiple testing using the Benjamini-Hochberg method. CpG probes with an adjusted p-value below 0.15 and having a difference in β-values of at least 0.05 (i.e. 5% difference in DNA methylation) between atherosclerosis patients and healthy controls were denoted as significant, and named sig-DMPs. DMRs were identified using the DMRcate R package [19]. A region was called significant when Pmean-value was below 0.001 with a maximum methylation difference of at least 5% and containing at least five CpGs. Sig-DMPs were annotated using the HumanMethylation450 v1.2 manifest file. The freely available EpiExplorer tool was used to add further annotation including chromatin state segmentation and histone modifications [20]. Enrichment or depletion of sig-DMPs in a particular genomic region was determined using the Fisher's exact test. Commercial Metacore (https://portal.genego.com/) and Ingenuity (www.ingenuity.com/) software packages were used to identify significant pathway enrichment of gene associated DMRs.

The method of Houseman et al. [21] incorporated in the RnBeads package was used to estimate the cell type composition in blood. Reference cell types for granulocytes, CD4+ T-cells, CD8+ T-cells, B-cells, monocytes and NK-cells were obtained from the study of

Reinius et al. [22] using the FlowSorted.Blood.450k R package. The methylation profiles were processed (filtering and normalization) together with the atherosclerosis methylation dataset in the same way as described above. In total 50,000 CpG probes with the highest variance were used to identify the top 500 CpG probes associated with the cell types. The relative cell type contributions were compared between healthy individuals and atherosclerosis using a normal student t-test. One-way ANOVA and Bonferroni Post-hoc test was used to detect methylation differences between the blood cell types using the data from Reinius et al.

Replication in atherosclerotic plaque material methylation dataset GSE46401

Normalized Infinium 450k DNA methylation data of atherosclerotic plaque material were obtained from GEO dataset GSE46401. The dataset contains data from 15 donor-matched aorta healthy and plaque tissue and from 19 carotid plaque material. A paired two-tailed student t-test was performed to find DNA methylation differences in the 15 donor-matched samples and an unpaired two-tailed student t-test was performed to find DNA methylation differences between the carotid plaque tissue samples and the healthy aorta samples.

Epityper Sequenom MassARRAY

In silico cleavage was done by means of the RSeqMeth script in R to aid selection of an optimal primer set for the genomic region of interest [23]. MassARRAY primers for regions in the BRCA1 (chr17:41,277,701-41,278,776) and CRISP2 (chr6:49,680,757-49,682,289) genes (**Supplementary Table 1 and Supplementary Figure 1 and 2**) were designed using the Sequenom EpiDesigner online tool (www.epidesigner.com). Bisulfite converted DNA was used for the methylation analysis. PCR reactions were performed using the following reagents: 10x buffer (Qiagen®), 10 mM dNTP, 10 µM primer mix, 5 U/µl HotStarTaqTM polymerase (Qiagen®) and deionized water. Methylation percentages were calculated based on the ratio of the unmethylated versus methylated peaks. In addition, DNA methylation standards (0, 20, 40, 60, 80 and 100%) were used to control for amplification bias. The R computing environment was used for the correction of the obtained methylation data according to standard procedures [24]. Linear regression was performed to fit the obtained data points according to the predicted standard methylation values. The student T-test was used to calculate the significance of the methylation difference between healthy and atherosclerotic blood samples.

Pyrosequencing

1 µg gDNA from each sample was bisulfite converted using the EpiTect Fast bisulfite conversion kit (Qiagen, Hilden, Germany) according to manufacturer's instructions. 15 ng of bisulfite treated DNA was subsequently used in PCR amplification using the PyroMark PCR kit (Qiagen, Hilden, Germany). Reverse primers were biotinylated to get biotin-labelled PCR products. Finally, DNA sequences were pyrosequenced using the PyroMark Q24 Advanced instrument (Qiagen, Hilden, Germany). First, streptavidin-coated Sepharose beads (High Performance, GE Healthcare, Uppsala, Sweden) were used to immobilize the biotin-labelled PCR products. Subsequently, PCR products were captured by the PyroMark vacuum Q24 workstation, washed and denaturated. The single stranded PCR products were mixed and annealed with their corresponding sequencing primer. After the pyrosequencing run was finished, the results were analyzed using the PyroMark Q24 Advanced software (Qiagen, Hilden, Germany). Biotinylated-reverse, forward and sequencing primers were designed using the PyroMark assay design 2.0 software (Qiagen, Hilden, Germany) (**Supplementary Table 1**).

BRCA1, NBR2 and CRISP2 methylation-expression correlation using TCGA, BLUEPRINT and ENCODE data

450k Illumina methylation data and RNAseq gene expression data from The Cancer Genome Atlas (TCGA) were obtained using the TCGABiolinks R package. For every TCGA cancer type, samples with both DNA methylation and gene expression data were matched and Spearman's correlation coefficients were calculated for every CpG probe in BRCA1-NBR2 and CRISP2 DMR.

BLUEPRINT bisulfite-sequencing data and RNAseq data were obtained using the DeepBlueR R package. Only cell types with both gene expression and DNA methylation data were selected and for every cell type median values were calculated. Spearman's correlation coefficients were than calculated between DNA methylation and gene expression for every CpG site located in BRCA1-NBR2 and CRISP2 DMR.

ENCODE 450k Illumina methylation data and RNAseq gene expression data were obtained using the ENCODE data portal. Only data from cell lines with both gene expression and DNA methylation data were downloaded. Values of cell lines with multiple entries were aggregated into a single value using the median. Spearman's correlation coefficients were than calculated between DNA methylation and gene expression for every CpG probe located in BRCA1-NBR2 and CRISP2 DMR.

Post-hoc analysis of human blood DNA methylation in BRCA1 and CRISP2 and subclinical atherosclerosis in middle-age healthy men

The AWHS is a study designed to assess cardiovascular risk and subclinical atherosclerosis in a cohort of middle-aged healthy men from Spain. The AWHS design and baseline characteristics have been reported elsewhere [2, 16]. In brief, in the baseline examination (2009-2010), the average (SD) age, body mass index, and waist circumference were 49.3 (8.7) years, 27.7 (3.6) kg/m2 and 97.2 (9.9) cm, respectively. The prevalence of overweight, obesity, current smoking, hypertension, hypercholesterolemia, and diabetes were 55.0, 23.1, 37.1, 40.3, 75.0, and 7.4%, respectively [25]. The adherence of the AWHS participants to the Mediterranean diet has been extensively studied [26]. 21.7% of participants in the AWHS reported being physical active (e,g, > 150 min/week or 30 min/d of jogging, walking quickly, dance, aerobics, gardening) [26]. The levels of physical activity were positively associated with the adherence to the Mediterranean lifestyle [26]. In 2011-2013, calcium coronary scoring was performed using non-contrast ECG gated prospective acquisition by a 16 multidetector computed tomography scanner (Mx 8000 IDT 16, Philips Medical Systems, Best, the Netherlands). During a single breath hold, images were acquired from the tracheal bifurcation to below the base of the heart. Scan parameters were 8 x 3 mm collimation, 220-mm field of view, 120 kVp, 55 mA, and 3-mm section thickness. Coronary calcium was quantified with calcium scoring software (Workspace CT viewer, Philips Medical Systems) that follows the Agatston method [27]. Carotid intima-media thickness was determined using the Philips IU22 ultrasound system (Philips Healthcare, Bothell, Washington). Ultrasound images were acquired with linear high-frequency 2-dimensional probes (Philips Transducer L9-3, Philips Healthcare), following the Bioimage Study protocol74. Examination of the carotid territory included the terminal portion (10 mm) of the common carotid, the bulb, and the initial portion (10 mm) of the internal and external carotid arteries. The given value for carotid artery intima-media thickness is the mean value from all sites at both sides. The AWHS study was approved by the Ethics Committee for Clinical Research at the Institutional Review Board of Aragón (CEICA) [3, 16]. All study participants provided written informed consent. The methods for DNA isolation and bisulfite conversion were similar to the methods implemented in the LVAD samples, which are standard manufacturer procedures. DNA methylation was measured using the platform Illumina Infinium Methylation 450K in a subsample of 23 individuals with available measures of subclinical atherosclerosis. Preprocessing and analysis of the Infinium 450k data was performed using the R package minfi [28]. CpG probes with a detection p-value higher than 0.01 were removed. Intra-array normalization was done using the Quantile Normalization. In exploratory analysis, we detected a potential batch effect by slide.

Methylation proportion values were represented as β-values ranging from 0 to 1. β-values were converted into M-values before doing the statistical analysis. For analysis of site-specific DNA methylation (independent variable) and subclinical atherosclerosis measures (dependent variable) in the AWHS, we estimated the differences in coronary artery score and carotid intima media thickness comparing 75th versus 25th percentiles of DNA methylation distribution at a given CpG site by linear regression with the following adjustment variables: age, smoking status (never, former and current smoking), body mass index, and houseman cell estimates (B cell, CD4+ and CD8+ T cells, granulocytes, monocytes and natural killer cells). Due to the small sample size we tried to avoid non-parsimonious regression parameters by performing two-stage regression for adjustment. First we adjusted DNA methylation M-values for potential confounders using combat [29] to correct for batch effect by slide. Subsequently, we adjusted intima media thickness and coronary artery calcium score levels for the same set of potential confounders. Second, we ran the final regression models using the residuals resulting from the first step recalibrated to the corresponding marginal mean. Since this was post-hoc analysis we evaluated the association of DNA methylation and subclinical atherosclerosis in CpG sites from regions validated in previous analysis (i.e. BRCA1 and CRISP2). Thus, we considered the non-Bonferroni corrected p-values <0.05 as statistically significant.

Results

Peripheral blood of atherosclerosis patients reveals no statistically significant changes in global DNA methylation in comparison to healthy individuals

Significant differences in clinical parameters were observed between the two study groups. C-reactive protein (CRP), hemoglobin concentration and triglycerides were significantly lower in the healthy individuals as compared to atherosclerosis patients (**Table 1**). The atherosclerosis patients were generally older than the healthy individuals.

DNA methylation profiles covering >450,000 CpG dinucleotides of peripheral blood of eight atherosclerosis and eight healthy individuals were generated by Illumina 450k BeadChip arrays. CpG probes were sub-grouped in relation to gene regions (TSS, gene body, 3'UTR, etc) and to CGIs to obtain the most comprehensive view of the DNA methylation distribution in both groups. Global DNA methylation was assessed in each individual by calculating the median beta-value of all CpG probes. The mean median

values were calculated per sample group (healthy and atherosclerosis). Overall, no statistical significant difference (P-value=0.9159) was observed in mean global DNA methylation between atherosclerosis patients (0.6613, SEM: 0.0067) versus healthy controls (0.6624, SEM: 0.0071). In addition, no global DNA methylation shifts were found between the groups when mapping CpG probes to different genomic locations (e.g. CGIs, shores, shelves).

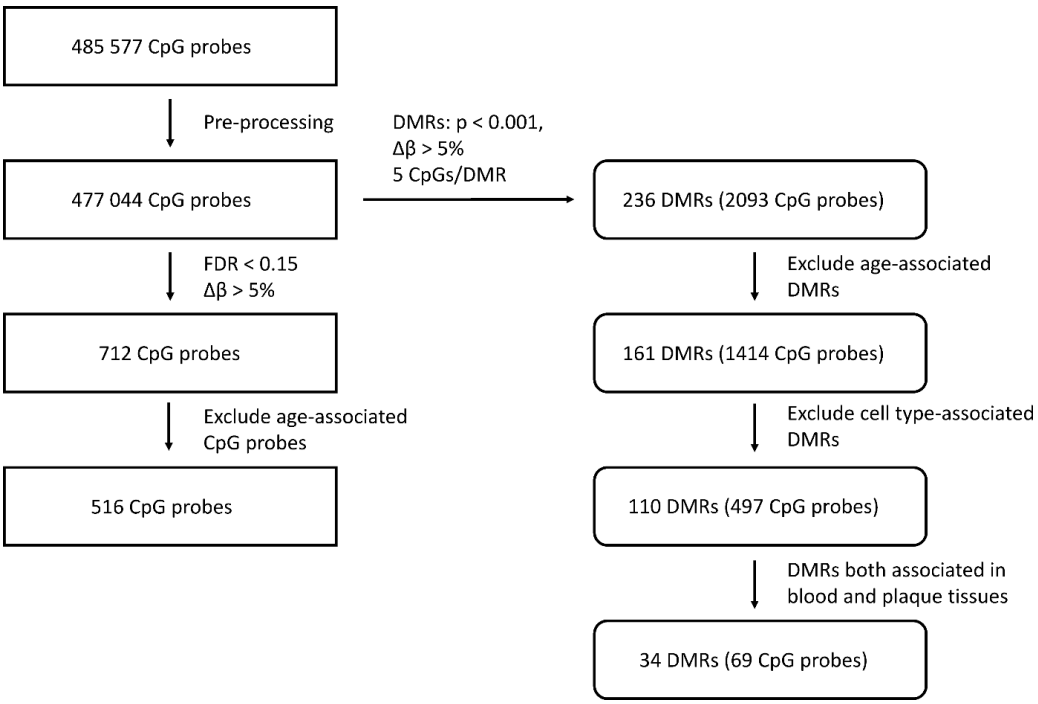


Figure 1: Flowchart of DMP and DMR selection.

Peripheral blood of atherosclerosis patients reveals specific DNA methylation signature

Criteria for the identification of atherosclerosis-DMPs (athDMPs) are summarized in Figure 1. Normalization, quality control and probe filtering for SNPs resulted in an output of 477,044 CpG probes. Differentially methylated CpG probes were filtered for a Benjamini-Hochberg adjusted p-value smaller than 0.15 and a difference in β-values between atherosclerosis patients and healthy controls of at least 0.05 (i.e. 5% difference in DNA methylation). 712 CpG probes met the selection criteria comprising 465 hypomethylated CpG sites (hypo-athDMPs) and 247 hypermethylated CpG sites (hyper-athDMPs) (**Figure 2A and Supplementary Table 2**). We observed a maximum of 20% difference in DNA methylation between atherosclerosis patients and healthy controls. Based on the methylation values of these CpG sites, principle component

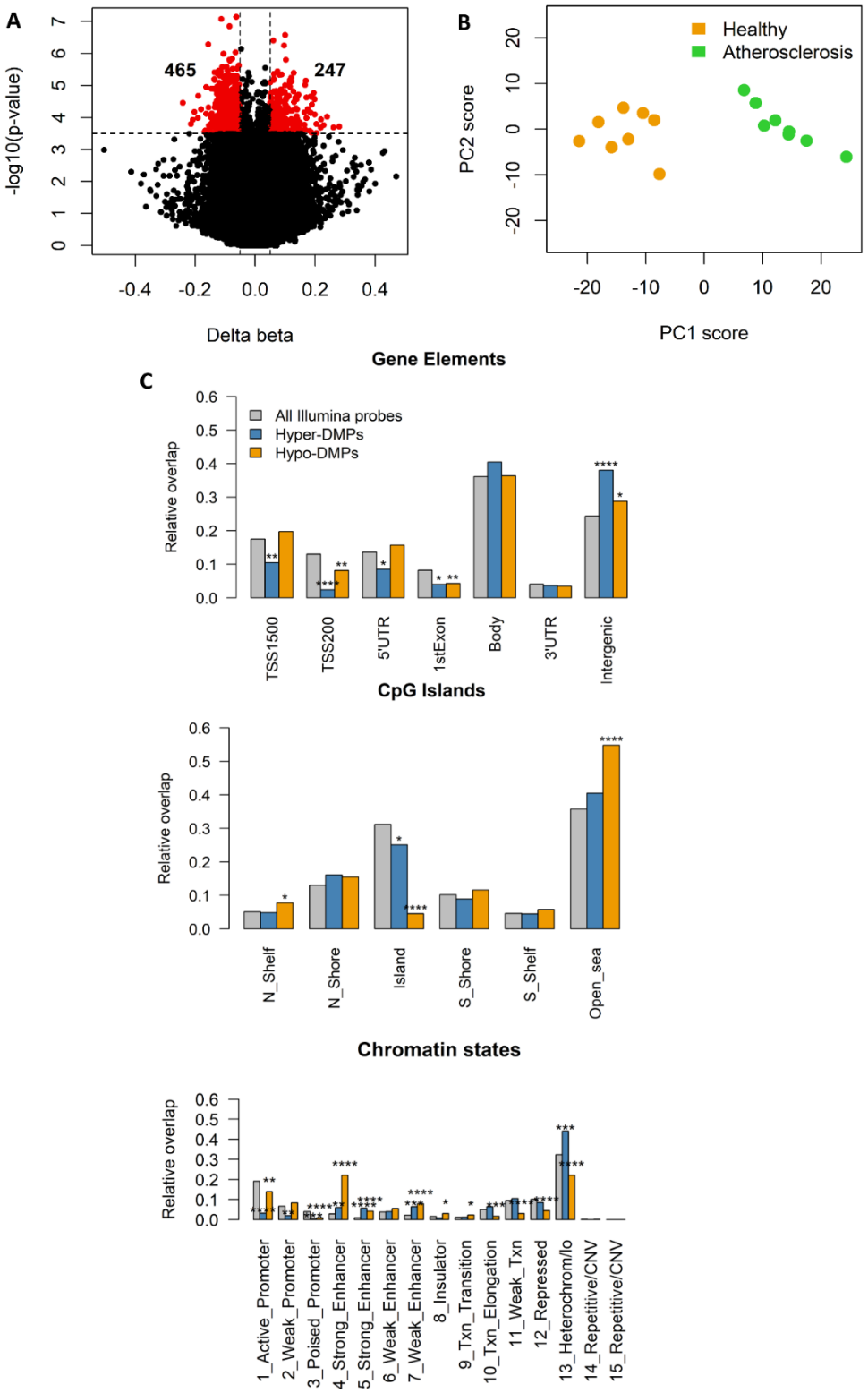

A
-log10(p-value)
465
247
Delta beta
B
Healthy
Atherosclerosis
PC2 score
PC1 score
C
Gene Elements
All Illumina probes
Hyper-DMPs
Hypo-DMPs
Relative overlap
TSS1500
TSS200
5'UTR
1stExon
Body
3'UTR
Intergenic
CpG Islands
N_Shelf
N_Shore
Island
S_Shore
S_Shelf
Open_sea
Chromatin states
1_Active_Promoter
2_Weak_Promoter
3_Poised_Promoter
4_Strong_Enhancer
5_Strong_Enhancer
6_Weak_Enhancer
7_Weak_Enhancer
8_Insulator
9_Txn_Transition
10_Txn_Elongation
11_Weak_Txn
12_Repressed
13_Heterochrom/lo
14_Repetitive/CNV
15_Repetitive/CNV

Figure 2: Differentially methylated CpG probes between healthy and atherosclerotic individuals. Sig-DMPs were selected based on FDR < 0.15 and delta beta > 5%. **A)** Volcano plot showing 465 hypo- and 247 hypermethylated probes meeting the selection criteria. **B)** PCA plot demonstrating the separation of atherosclerosis patients for CVD (green) and healthy individuals (orange) based on the DNA methylation values of the 712 DMPs. **C)** Genomic distribution of sig-DMPs based on gene regions, CpG-island regions and chromatin segmentation states (based on GM12878 cell type data). Significant enrichment or depletion of sig-DMPs (P-value < 0.05), determined by the Fisher's exact test, are marked by an asterisk.

analysis reveals a clear separation of DNA methylation profiles of atherosclerosis and healthy individuals (**Figure 2B**). Analysis of functional genomic locations revealed that both hyper- and hypo-athDMPs were depleted in promoter regions and CGIs and enriched in intergenic, CpG-poor and enhancer regions (**Figure 2C**).

Since median age of atherosclerosis patients (78±8 years old) and healthy controls (47±8 years old) was significantly different, we next overlapped our sig-DMP list with the list of age-responsive CpG probes, identified by Steengenga et al. [30], to discriminate between aging- and atherosclerosis-specific DNA methylation changes. Of the 7,477 CpG probes with age-dependent methylation changes, 196 probes overlapped with our athDMP list, resulting in 516 age-independent DMPs, including 287 hypo- and 229 hyper-DMPs (**Supplementary Table 2**).

Blood samples of atherosclerosis patients reveal a different immune cell type composition in comparison to healthy individuals

Since blood is a heterogeneous collection of different cell types, each characterized by unique DNA methylation profile, as recently demonstrated by Jaffe and colleagues [31], we next wanted to evaluate whether identified DNA methylation changes did not simply reflect variation in blood cell composition between both studied populations. Using the algorithm developed by Houseman and colleagues [21] for mathematical deconvolution of relative immune cell type composition of blood samples based on Illumina 450k data [22], we observed that the fraction of granulocytes was slightly but statistical significantly (P<0.05) increased in atherosclerosis patients in comparison to control individuals (**Figure 3A** and **Supplementary Table 4**). In contrast, the CD8+ T-cell population was significantly reduced in atherosclerosis patients. Although statistically not significant, a fraction of CD4+ T- and NK-cells tended to decrease in atherosclerosis patient. Because of the relative low sample size analyzed, we were not able to correct for this in the linear model analysis.

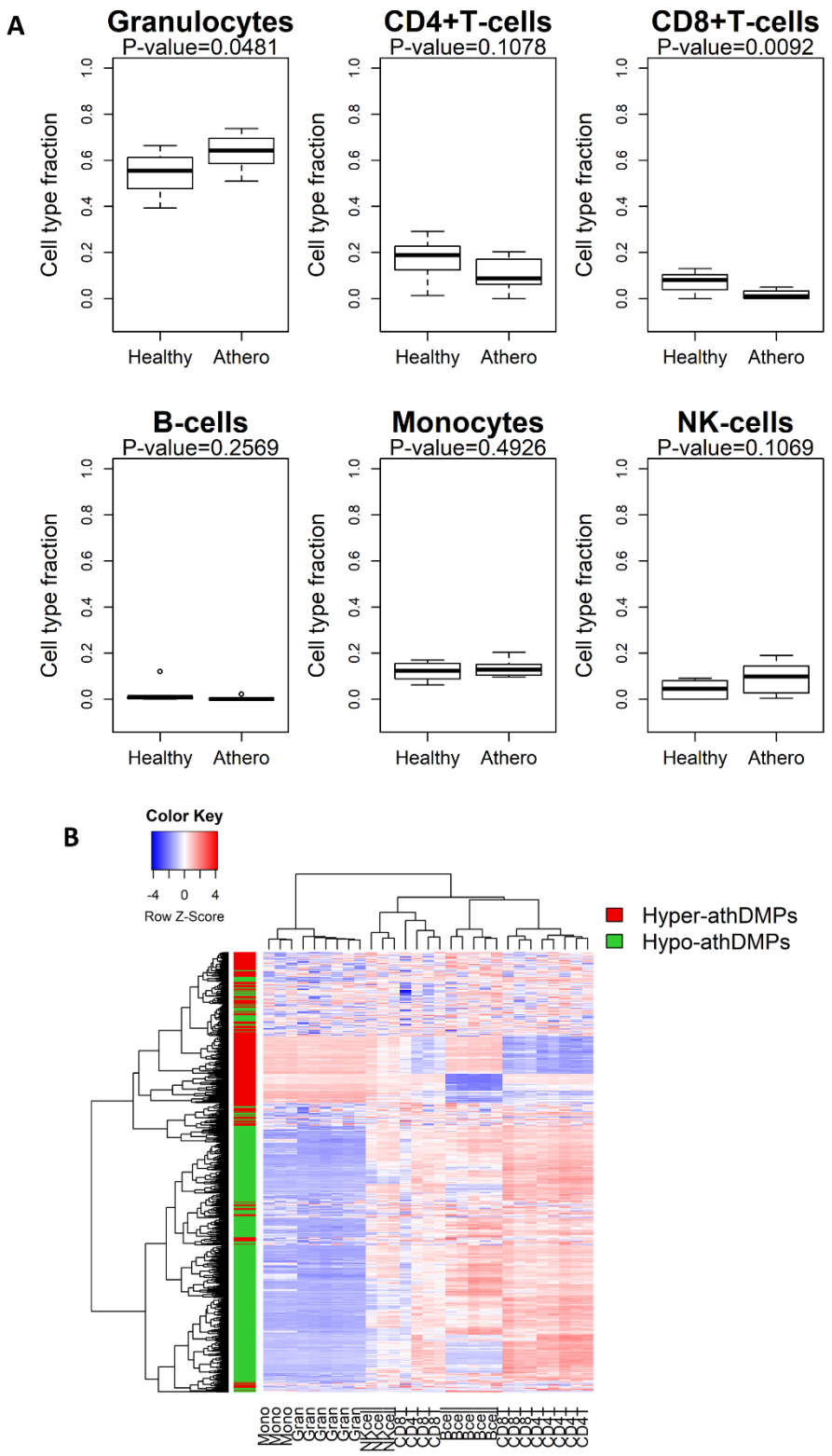

A
Granulocytes
P-value=0.0481
CD4+T-cells
P-value=0.1078
CD8+T-cells
P-value=0.0092
B-cells
P-value=0.2569
Monocytes
P-value=0.4926
NK-cells
P-value=0.1069
Cell type fraction
Healthy
Athero
B
Color Key
-4 0 4
Row Z-Score
Hyper-athDMPs
Hypo-athDMPs

Figure 3: Deconvolution of immune cell type blood composition. The approach described by Houseman et al. was applied to determine the relative immune cell type fraction (Y-axis) in healthy and atherosclerosis blood samples. **A)** Statistical significant differences in cell type contribution between healthy and atherosclerosis blood samples were calculated by a student t-test. **B)** DNA methylation values of the 712 DMPs in immune cell types. The hyper-DMPs are colored in red and the hypo-DMPs in green.

We also determined the methylation status of the athDMPs among the different blood cell types, and noticed that a large fraction of these CpG probes differed (**Figure 3B**). Hyper-athDMPs were mainly higher in methylation in granulocytes and monocytes, while hypo-athDMPs were predominantly lower in methylation in granulocytes and monocytes compared to T-, B- and NK-cells. However, about 20% of the athDMPs cannot be attributed by cell type heterogeneity.

Pathway enrichment analysis of common gene associated DMRs in blood, aorta and carotid plaque material reveals impaired NRF2 oxidative stress, DNA repair, thioredoxin and inflammatory pathways

To further reduce biological complexity of DNA methylation changes, we next determined consecutive athDMPs using the R-package DMRcate. In total 236 sig-DMRs were identified (Pmean-value < 0.001) containing at least 5 consecutive CpG probes with a minimal DNA methylation difference of 5% (Δβ-value > 0.05) (**Figure 1 and Supplementary Table 3**). After overlap with known age-dependent CpG probes, 75 DMRs were removed leaving a total of 161 DMRs (athDMRs), containing 1,424 CpG probes, for further analysis.

In total, 51 cell type-associated DMRs were excluded using the houseman method leaving 110 athDMRs for further analysis. To replicate the potential role for the 110 athDMRs as blood-based surrogate biomarkers for plaque biopsy material in atherosclerosis, we compared our findings with publicly available data from Zaina et al (GSE46401) [32] which contains DNA methylation profiles from donor-matched healthy and atherosclerotic human aorta tissue and from human carotid plaque samples in a larger independent cohort [32]. We performed a two-tailed student t-test for the 497 CpG probes located in the 110 sig-DMRs comparing methylation values between donor-matched healthy and aorta plaque tissue and between healthy aorta and carotid plaque tissue. We found that 69 CpG probes located in 34 unique DMRs were both consistently differentially methylated in our blood-based dataset, in aorta plaque tissue and carotid plaque tissue (**Supplementary Table 5**). To find out the relationship between the different sig-DMRs, we mapped each sig-DMR to the nearest gene and constructed a network using the GeneMANIA Cytoscape plugin (**Figure 4A**). Interestingly the breast cancer 1 gene (BRCA1) stands out as one of the most highly connected node in our network, physically interacting with nine other proteins. In addition, cysteine rich secretory protein 2 (CRISP2), was the gene with the highest node degree. Due to their

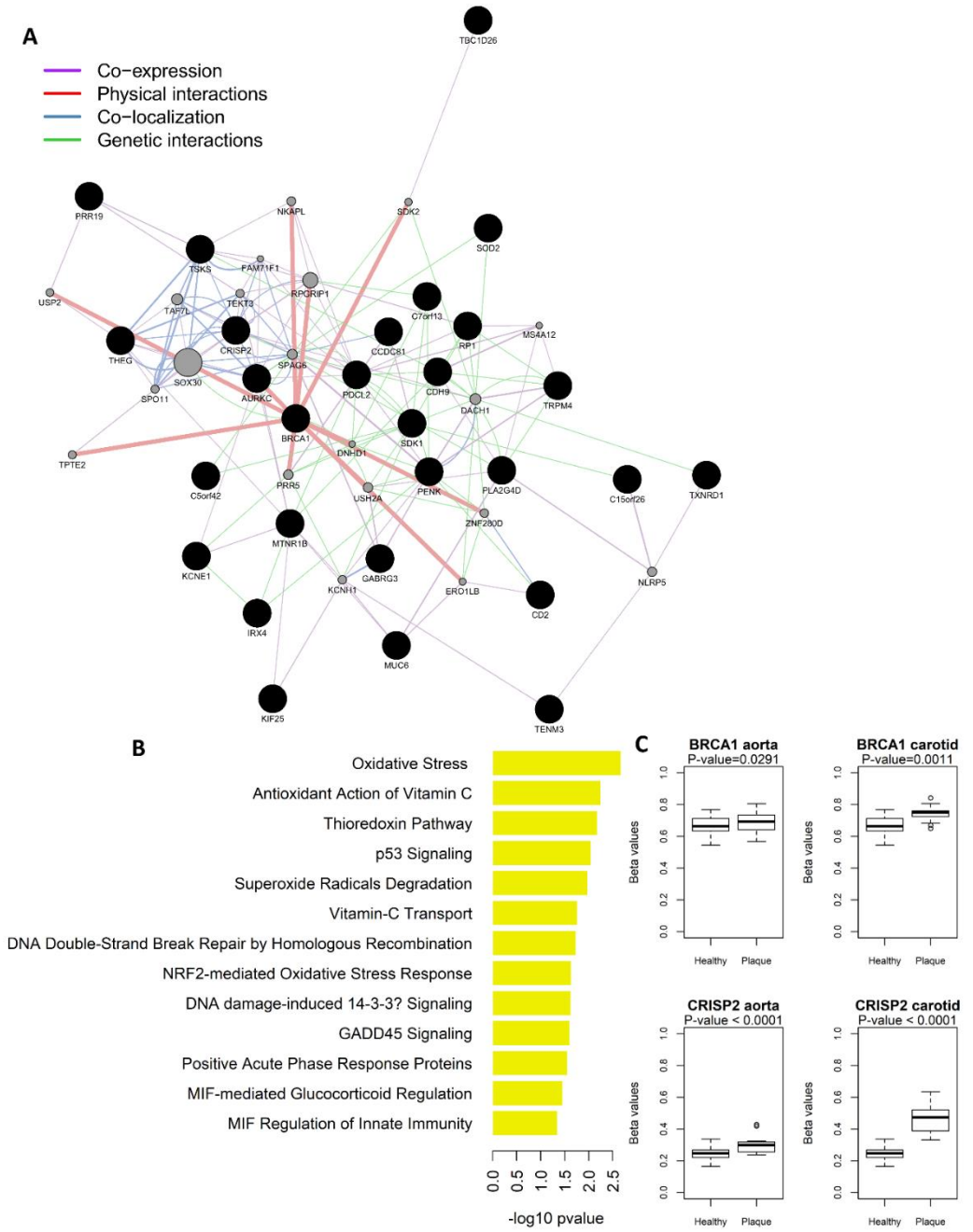


Figure 4: Differentially methylated regions (sig-DMRs) not affected by age or cell type composition and consistently differentially methylated in both blood and plaque. A) GeneMANIA network of the 34 sig-DMRs. **B)** IPA based pathway enrichment analysis of gene associated DMRs in atherosclerosis. **C)** DMRs associated with BRCA1and CRISP2 in an independent publicly available cohort [32]. The Zaina et al. cohort [32] comprises of 30 donor-matched atherosclerotic plaque tissue samples and 19 carotid plaque samples.

high network interconnectivity (high node degree), BRCA1 and CRISP2 genes were selected for further DNA methylation biomarker validation studies. Both genes were found to be hypermethylated in atherosclerotic whole blood and plaque samples (**Figure 4B**). Interestingly, pathway enrichment analysis of the 34-common gene associated DMRs revealed Nuclear factor (erythroid-derived 2)-like 2 (NRF2), oxidative stress (SOD2), DNA repair (BRCA1), thioredoxin (TXNRD1) and inflammatory pathways (MIF, PLA2G4D) (**Figure 4C and Supplementary Table 6**).

Verification of Illumina 450K DNA methylation intensities of BRCA1-NBR2 and CRISP2 DMRs by Epityper MassARRAY

DMRs of BRCA1 and CRISP2 genes determined by Illumina 450 array were selected for further technical validation by Epityper MassARRAY. For BRCA1, we decided to focus on 14 consecutive hypermethylated CpG probes located in a CGI located at the promoter site of BRCA1 and NBR2 (chr17:41,277,974-41,278,445, **Supplementary Figure 1**). Nine CpG probes in this region showed more than 10% DNA hypermethylation in the atherosclerosis patients (cg26370022, cg15065591, cg02286533, cg18372208, cg14947218, cg16006004, cg06001716, cg25288140 and cg24900425). Finally, the promoter region of the CRISP2 was selected as a second amplicon for validation (**Supplementary Figure 2**). Seven neighboring CpG sites (cg26715042, cg14997592, cg04595372, cg01706515, cg25390787, cg08942800 and cg01076129) were found to be more than 10% hypermethylated in this region for the atherosclerosis patients. Altogether, the Epityper MassARRAY and Illumina DNA methylation levels revealed strongly significant correlations for BRCA1 (ρ = 0.711 – 0.932) and CRISP2 (ρ = 0.680) (**Supplementary Figure 3A**). Moreover, BRCA1-NBR2 and CRISP2 target genes demonstrate significant hypermethylation in atherosclerosis blood samples, as compared to blood samples derived from healthy individuals (**Supplementary Figure 3B**). Similar, DNA atherosclerosis associated BRCA1 DNA hypermethylation results were also obtained by pyrosequencing of the BRCA1 DMR chr17:41278125-41278228 (including cg26279233, cg cg06001716 and cg cg14947218), whereas no valid pyrosequencing assay could be designed for the CRISP2 DMR (**Supplementary Figure 3C**).

We further compared the methylation status of the blood cells types in the two validated regions using the reference methylation data set of Reinius et al [22]. Four CpG probes in the BRCA1 DMR (cg26370022, cg11529738, 14947218 and cg06001716) showed significant DNA methylation differences between immune cell types ($p<0.05$, not corrected for multiple testing) (**Supplementary Figure 4**). However, after Bonferroni correction hypermethylation of only one CpG probe remained significantly correlated with immune cell type (cg26370022). In the CRISP2 DMR significant immune

cell type specific DNA methylation changes were observed for two CpG probes (cg01706515 and cg21710255). However, observed atherosclerosis related DNA hypermethylation trend for most BRCA1 and CRISP2 CpG probes does not follow the expected methylation change in blood samples enriched for granulocyte and reduced CD8+T immune cell subpopulations. As such, our results show that DNA hypermethylation of most BRCA1-NBR2 and CRISP2 CpG probes occurs independently of sample variation in blood cell type composition and is associated with atherosclerosis pathology.

To determine whether methylation at the BRCA1-NBR2 and CRISP2 promoter site regulates gene expression, we used publicly available data from TCGA, and correlated the Illumina methylation with RNAseq gene expression data. We found both positive as negative correlations between DNA methylation at BRCA1-NBR2 promoter and BRCA1 gene expression values (**Figure 5**). BRCA1 shares his promoter with the NBR2 non-coding RNA which is transcribed in the reverse direction. In almost all cancer types, BRCA1-NBR2 promoter methylation was negatively correlated with NBR2 gene expression. Also methylation at CRISP2 promoter region was negatively correlated with CRISP2 gene expression in most of the TCGA cancers. A similar analysis was performed using RNAseq and BSseq data from different blood cells obtained from the BLUEPRINT database, and using RNAseq and 450k Illumina data from ENCODE cell lines (**Supplementary Figure 5**). In the BLUEPRINT and ENCODE datasets, the correlations were much weaker. Using the BLUEPRINT samples, we observed slight negative correlation with BRCA1 gene expression and slight positive correlations with NBR2 gene expression. However, this was not observed using the ENCODE cell lines. Finally, CRISP2 gene expression was low or absent in almost all BLUEPRINT and ENCODE samples (**Supplementary Figure 5A**) .

Altogether, our data suggest that BRCA1-NBR2 and CRISP2 DNA methylation patterns are potential epigenetic biomarkers related to atherosclerosis in blood and atherosclerotic plaque tissue matrix.

Association of blood DNA methylation changes in BRCA1-NBR2 and CRISP2 with subclinical atherosclerosis in healthy middle age men of the AWHS cohort

To explore the role of validated regions as predictors of subclinical atherosclerosis in a population with a low burden of disease, we evaluated the association of DNA methylation levels from fresh frozen whole blood samples collected at the baseline visit (2009-2010) and subclinical atherosclerosis measured in 2011-2013 from a subsample of 24 AWHS participants with available baseline InfiniumMethylation450K data. We

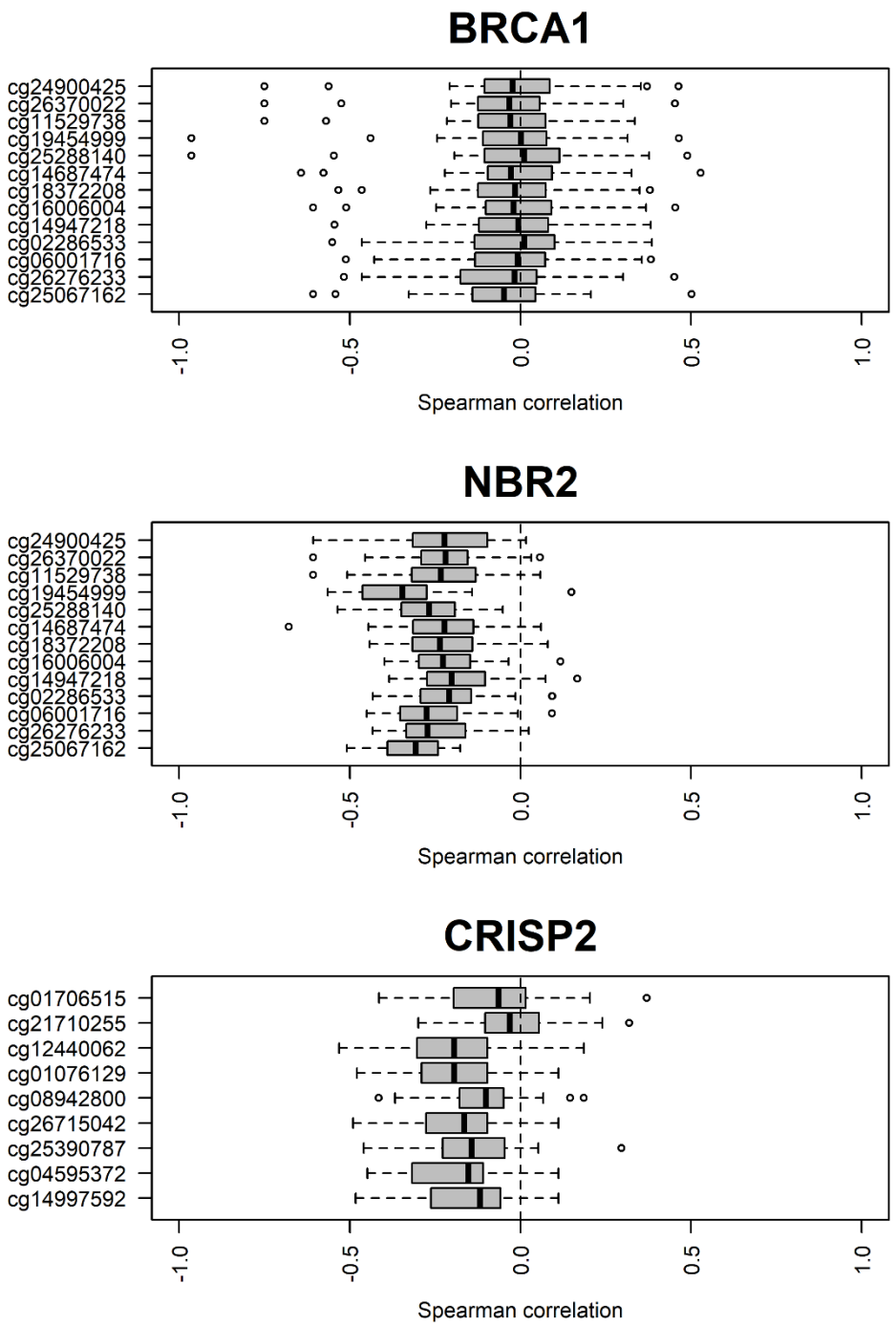


Figure 5: Correlation between gene expression and DNA methylation in TCGA cancers. Boxplots represent Spearman's correlation coefficients of the correlation between DNA methylation and gene expression in BRCA1, NBR2 and CRISP2 for each TCGA cancer.

looked for associations between blood DNA methylation and the subclinical atherosclerosis measures, coronary calcium score and carotid intima media thickness. Associations were found for three CpG probes located in CRISP2 (cg12440062, cg25390787, cg01076129) and one CpG probe in BRCA1-NBR2 (cg16630982). The strongest statistically significant CpG probe was cg12440062 (located 70 bases upstream of the promotor) for CRISP2, consistently, for both coronary calcium score and carotid intima media thickness measures (**Table 2**). The multi-adjusted difference in coronary calcium score comparing the 75th to the 25th percentiles of DNA methylation was -46.62 score points (-86.87, -6.36; p-value = 0.03) for cg12440062 in CRISP2. The corresponding difference in carotid intima media thickness was -0.20 millimeters (-0.33, -0.06; p-value = 0.009) for cg12440062 in CRIPS2. For CRISP2, DNA methylation in cg01076129 (located in promoter region) was associated with carotid intima-media thickness, but not coronary calcium score. The association with cg25390787 (located in promotor region) was only borderline significant (p=0.06) consistently for both atherosclerosis measures. With respect to the BRCA1-NBR2 DMR, the strongest association with both coronary calcium score (p-value = 0.018) and carotid intima media thickness (p-value = 0.0019) was found for cg16630982 (located in the promotor region), whereas other cg probes did not reach significance within the limited sample series tested.

Table 2: Significant associations of BRCA1-NBR2 and CRISP2 CpG probes with coronary calcium score and carotid intima thickness.

Coronary calcium score			
	Gene	**Difference P75vsP25 (95% CI)**	**P-value**
cg16630982	BRCA1-NBR2	-36.59 (-64.53 - -8.64)	0.018
cg12440062	CRISP2	-46.65 (-86.87 - -6.36)	0.03
cg25390787	CRISP2	34.98 (-69.35 - -0.61)	0.06
Carotid intima thickness			
	Gene	**Difference P75vsP25 (95% CI)**	**P-value**
cg16630982	BRCA1-NBR2	-0.16 (-0.25 - -0.07)	0.0019
cg12440062	CRISP2	-0.20 (-0.33 - -0.06)	0.009
cg01076129	CRISP2	-0.14 (-0.26 - -0.02)	0.036
cg25390787	CRISP2	-0.12 (-0.24 - 0.00)	0.06

Discussion

In this study, we identified genomic regions in BRCA1-NBR2 and CRISP2 which were consistently differentially methylated in blood DNA of atherosclerosis patients

compared to healthy individuals, and in aortic and carotid plaque samples compared to aorta samples without plaque. Furthermore, methylation in BRCA1-NBR2 and CRISP2 DMR was also associated with subclinical atherosclerosis measures in an independent sample of middle age men. Our results thus support a potential role of blood DNA surrogate markers for early CVD detection.

Epigenetics may provide the missing mechanism linking environment, genome and atherosclerotic phenotype. Identifying epigenomic biomarkers that parallel the development of subclinical atherosclerosis might open new paths for risk stratification and prevention, and may help to further understand the pathophysiology of atherosclerosis. In particular, changes in DNA methylation patterns have been linked to several cardiovascular-related biomarkers, including homocysteine and CRP [15]. Furthermore, an increasing number of studies report DNA methylation alterations in atherosclerosis [33-35]. For example, a recent genome-wide study showed DNA methylation differences between healthy donor-matched aortic healthy and plaque tissue, indicated by epigenetic drift of DNA methylation in aortic plaques with atherosclerotic progression [32, 36]. Furthermore, known cardiovascular risk factors, including homocysteine levels, smoking and age have been described to induce DNA methylation changes [37-39]. Currently, only few DNA methylation studies have been performed with blood samples of CAD patients. Sharma et al. identified 72 hypermethylated DMRs associated with CAD and hyperhomocysteinemia using a 12k human CGI microarray [40]. Guay et al. performed a study on subjects with familial hypercholesterolemia with or without CAD using the Infinium 27k methylation array [41]. Since blood leukocytes have major contributions to the initiation, progression and maintenance of atherosclerosis, we determined genome-wide DNA methylation profiles in blood samples of atherosclerosis patients, in comparison to healthy individuals to identify CVD related epigenetic biomarkers. Although Illumina 450K profiling did not reveal significant global DNA methylation differences between atherosclerosis patients and healthy individuals, HPLC based methods demonstrated global DNA hypermethylation in blood leukocytes [13, 15] whereas both global DNA hypermethylation and hypomethylation have been reported in atherosclerotic vascular tissue [32, 42, 43].

Upon further mapping of DNA methylation changes at specific CpG motifs or regions, 161 DMRs were identified to be differentially methylated based on specific selection criteria. Of particular interest, pathway enrichment analysis of gene associated DMRs revealed DNA impaired epigenetic regulation of integrin and cadherin dependent cell adhesion, cell cycle, cell death, chemotaxis, immune phagocytosis and estrogen hormone pathways which are all critically involved in atherosclerosis. In accordance with other studies examining DNA methylation in metabolic diseases, the observed

methylation changes were relatively small (max 20%), as compared to cancer specific DNA methylation changes. More specifically, whereas promoter regions and CGIs were found to be depleted of atherosclerosis associated DMPs, gene bodies, intergenic and open sea regions show enrichment of various DMPs. Interestingly, a strong enrichment was also observed in enhancer regions suggesting that impaired control of distal regulatory regions may contribute to atherosclerosis.

Remarkably, mathematical deconvolution (Houseman correction) of the blood DNA methylation profiles revealed significant changes in the granulocyte and CD8+ T immune cell populations in atherosclerosis patients as compared to healthy individuals, which could be highly relevant for atherosclerotic plaque formation. More particularly, important regulatory roles for granulocyte and CD4+/CD8+ T cell populations have recently been identified in atherosclerotic lesions and coronary thrombus evolution [44, 45]. As expected, a large fraction of the sig-DMRs were also differentially methylated between blood cell types, suggesting that their change in DNA methylation in atherosclerosis patients could be due to a difference in blood cell type composition between atherosclerosis and healthy controls. Heterogeneity of blood samples could be prevented using cell count and sorting methods (fluorescence-activated cell sorting, FACS) to analyze specific immune cell subpopulations. However, these methods are difficult and costly to apply in large epidemiological cohort studies. CpG sites associated with blood cell type were excluded for further analysis, leaving 110 sig-DMRs not affected by blood cell types.

Interestingly, of the 110 remaining DMRs (comprising 497 CpG sites), 34 DMRs (69 CpG sites) were also found to be differentially methylated in plaque tissues, which suggests that blood-associated epigenetic biomarkers can be valid surrogate markers for methylation changes in plaque material. Of particular interest, pathway enrichment analysis of the 34 common gene associated DMRs revealed epigenetic impairment of NRF2 oxidative stress (SOD2), DNA repair (BRCA1), thioredoxin (TXNRD1) and inflammatory pathways (MIF, PLA2G4D) in atherosclerosis conditions. Upon further network analysis of each sig-DMR, mapped to the nearest gene, we identified a highly interconnective network with central roles of BRCA1 and CRISP2 DMRs, which prompted us to focus on these genes for further validation. Of special note, the DMRs in BRCA1 and CRISP2 appeared to be largely independent of age and/or immune cell type composition, and both hold promise as valuable atherosclerosis related biomarkers in routine blood analysis. Interestingly, DNA methylation at the BRCA1 locus was already present in healthy controls, which is contrary to other studies where a lack of DNA methylation was observed [46, 47]. Nevertheless, it must be emphasized that the methylated CpGs in our data set are located in a CGI approximately 600 bp upstream of the BRCA1 gene, whereas previous studies [47] detected hypomethylation

around the transcription start site of the gene, which can also be appreciated in our data (**Supplementary Figure 1**).

DNA methylation changes of the BRCA1 and CRISP2 DMRs could also be replicated in an Illumina 450K dataset of paired atherosclerotic plaque and normal aorta samples from 24 middle aged men with subclinical atherosclerosis of the AWHS. Moreover, we also observed statistically significant association between DNA methylation in several CpG sites in these regions and coronary calcium score and carotid intima-media thickness when using data from this well-established population-based cohort [3, 16], thus adding robustness to our findings. Surprisingly, the associations with subclinical atherosclerosis measures were not always directionally consistent compared to the associations comparing blood DNA methylation of CVD versus healthy individuals or the atherosclerosis versus normal aorta samples. The explanation for this inconsistency still remains unclear, although we cannot exclude cell type specific variations in blood sample composition or complex genotype SNP, microRNA or lncRNA dependent heterogenic epigenetic regulation of different BRCA1 or CRISP2 variants [48-54]. In addition, changes in lifestyle (diet, smoking, pollution, exercise) and pharmacological treatments (statins, aspirin, PARP inhibitors) could further obscure DNA methylation changes associated with atherosclerosis [48, 55].

However, the known biological role of BRCA1 and CRISP2 in cardiometabolic risk and inflammation pathways adds further significance to our findings. Besides the well described tumor suppressor function of BRCA1 in breast and ovarian cancers, more recent research also demonstrates an important role for BRCA1 in suppression of endothelial dysfunction and atherosclerosis [56]. In the latter study, BRCA1-overexpressing ApoE knock-out mice developed significantly less atherosclerotic plaque lesions together with reduced macrophage infiltration and diminished ROS production [56]. In another study, women lacking functional BRCA1/BRCA2 at increased breast cancer risk also show greater risk for heart disease and metabolic diseases [57-60]. BRCA1 is involved in multiple cellular processes and genome stability maintenance like DNA repair, transcriptional regulation, ubiquitination and cell-cycle control [61]. Excessive production of reactive oxygen species, in part via upregulation of DNA damage pathways, is a central mechanism governing pathologic activation of vascular smooth muscle cells. Remarkably, BRCA1 was found to shield vascular smooth muscle cells (VSMCs) from oxidative stress by inhibiting NADPH Nox1-dependent reactive oxygen species production [62]. More recently, BRCA1 was found to regulate lipogenesis through its interaction with acetyl coenzyme A carboxylase [58]. Along the same line, BRCA1 plays a critical role in the regulation of metabolic function in the skeletal muscle where it is involved in lipid storage and insulin resistance [63]. In analogy to BRCA1 dependent suppression of cell motility and epithelial-mesenchymal

transition in cancer [64], BRCA1 may also prevent endothelial-mesenchymal transition involved in atherosclerosis progression and other CVDs (myocardial infarction, vascular calcification) [65-67]. As our data suggests, silencing of the BRCA1 gene in atherosclerosis patients may mean that this gene not only acts a tumor suppressor but also as a vascular protector against oxidative cell damage. Of note, using TCGA cancer data BRCA1 methylation-expression correlation seems to be complex, with both negative and positive correlations. In addition, in TCGA cancers, methylation seems to be stronger correlated with NBR2 gene expression. NBR2 is a lncRNA sharing a bi-directional promoter with BRCA1. A role for the lncRNA NBR2 in CVD has not yet been described, but it seems to function as a tumor suppressor by activating AMPK and regulating autophagy and metabolic energy stress pathways [68, 69]. Interestingly, AMPK plays an important role in atherosclerosis via autophagy regulation. Moreover, AMPK activation reduces the formation of atheromata-inducing macrophages [70, 71]. In contrast, non-cancer samples obtained from BLUEPRINT and ENCODE datasets reveal only weak or no correlations. These results suggest that BRCA1-NBR2 DMR has a complex disease related relationship with BRCA1 and NBR2 gene expression.

While no CRISP2 functions have so far been reported in relation to CVDs, CRISP2 gene activities were recently associated with oxidative stress responses and decline of lung function upon smoke or particular matter exposure [72]. In another study, CRISP expression was found to abolish the neovascularization process induced by exogenous growth factors (bFGF, vpVEGF) [55]. Decreased CRISP2 expression correlated with Th2-like eosinophilic inflammation in chronic nasal asthmatic chronic rhinosinusitis [73]. As such, the potential involvement of CRISP2 in CVD pathologies and angiogenesis via oxidative stress and inflammatory responses warrants further investigation. However, CRISP2 gene expression in blood cells and cell lines was low or even absent, questioning the functional relevance of this methylation change.

An important limitation in our study is reflected by the small sample size of the studied samples, which renders our analysis clearly underpowered. Future prospective studies in larger and distinct cohorts could further enable the validation of BRCA1 and CRISP2 to predict early CVD. Additionally, the atherosclerosis patients were older than the healthy controls. Even though, a correction for age-specific methylation was performed, we cannot exclude that age may affect the results and is therefore a confounding factor. In addition, as atherosclerosis is an age-dependent disease, probably some of the excluded age-dependent CpG sites overlap with CDV related CpGs. Nonetheless, in the post-hoc analysis with the AWHS, a study population composed of CVD free middle age men, the main findings were consistent even after adjustment of age, BMI, smoking and houseman cell composition. While we cannot discard a potential lack of generalizability, which is typical of observational studies, an

important strength of our study includes the availability of DNA-methylation data from three independent set of samples covering the whole spectrum from blood samples from individuals with subclinical atherosclerosis measures and individuals at high and low cardiovascular risk to carotid and aorta samples.

In conclusion, we identified promising novel epigenetic biomarkers of clinical and subclinical atherosclerotic disease, in genes involved in impaired leukocyte-endothelium functions during atherosclerosis progression. These regions deserve further consideration in experimental studies and prospective population-based cohort to confirm their potential role in cardiovascular risk. If confirmed, the reported markers could become potential tools to support early identification of individuals at high CVD risk who could benefit from preventive interventions for CVD prevention and control.

References

1. Kwon, G.P., et al., *Contribution of macromolecular structure to the retention of low-density lipoprotein at arterial branch points.* Circulation, 2008. **117**(22): p. 2919-27.
2. Fernandez-Ortiz, A., et al., *The Progression and Early detection of Subclinical Atherosclerosis (PESA) study: rationale and design.* Am Heart J, 2013. **166**(6): p. 990-8.
3. Fernandez-Friera, L., et al., *Prevalence, Vascular Distribution, and Multiterritorial Extent of Subclinical Atherosclerosis in a Middle-Aged Cohort: The PESA (Progression of Early Subclinical Atherosclerosis) Study.* Circulation, 2015. **131**(24): p. 2104-13.
4. Laclaustra, M., et al., *Femoral and Carotid Subclinical Atherosclerosis Association With Risk Factors and Coronary Calcium: The AWHS Study.* J Am Coll Cardiol, 2016. **67**(11): p. 1263-74.
5. Ginhoux, F. and S. Jung, *Monocytes and macrophages: developmental pathways and tissue homeostasis.* Nat Rev Immunol, 2014. **14**(6): p. 392-404.
6. Xue, J., et al., *Transcriptome-based network analysis reveals a spectrum model of human macrophage activation.* Immunity, 2014. **40**(2): p. 274-88.
7. Wang, J., et al., *Hyperhomocysteinemia-Induced Monocyte Chemoattractant Protein-1 Promoter DNA Methylation by Nuclear Factor-kappaB/DNA Methyltransferase 1 in Apolipoprotein E-Deficient Mice.* Biores Open Access, 2013. **2**(2): p. 118-27.
8. Baccarelli, A., et al., *Ischemic heart disease and stroke in relation to blood DNA methylation.* Epidemiology, 2010. **21**(6): p. 819-28.
9. Cash, H.L., et al., *Cardiovascular disease risk factors and DNA methylation at the LINE-1 repeat region in peripheral blood from Samoan Islanders.* Epigenetics, 2011. **6**(10): p. 1257-64.
10. Guay, S.P., et al., *ABCA1 gene promoter DNA methylation is associated with HDL particle profile and coronary artery disease in familial hypercholesterolemia.* Epigenetics, 2012. **7**(5): p. 464-72.
11. Huang, Y.S., Y.F. Zhi, and S.R. Wang, *Hypermethylation of estrogen receptor-alpha gene in atheromatosis patients and its correlation with homocysteine.* Pathophysiology, 2009. **16**(4): p. 259-65.
12. Jiang, D., et al., *Elevated PLA2G7 gene promoter methylation as a gender-specific marker of aging increases the risk of coronary heart disease in females.* PLoS One, 2013. **8**(3): p. e59752.
13. Kim, M., et al., *DNA methylation as a biomarker for cardiovascular disease risk.* PLoS One, 2010. **5**(3): p. e9692.
14. Levenson, V.V. and A.A. Melnikov, *DNA methylation as clinically useful biomarkers-light at the end of the tunnel.* Pharmaceuticals (Basel), 2012. **5**(1): p. 94-113.
15. Sharma, P., et al., *Detection of altered global DNA methylation in coronary artery disease patients.* DNA Cell Biol, 2008. **27**(7): p. 357-65.
16. Casasnovas, J.A., et al., *Aragon workers' health study--design and cohort description.* BMC Cardiovasc Disord, 2012. **12**: p. 45.
17. Assenov, Y., et al., *Comprehensive analysis of DNA methylation data with RnBeads.* Nat Methods, 2014. **11**(11): p. 1138-1140.
18. Teschendorff, A.E., et al., *A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data.* Bioinformatics, 2013. **29**(2): p. 189-96.
19. Peters, T.J., et al., *De novo identification of differentially methylated regions in the human genome.* Epigenetics Chromatin, 2015. **8**: p. 6.
20. Halachev, K., et al., *EpiExplorer: live exploration and global analysis of large epigenomic datasets.* Genome Biol, 2012. **13**(10): p. R96.
21. Houseman, E.A., et al., *DNA methylation arrays as surrogate measures of cell mixture distribution.* BMC Bioinformatics, 2012. **13**: p. 86.
22. Reinius, L.E., et al., *Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility.* PLoS One, 2012. **7**(7): p. e41361.
23. Coolen, M.W., et al., *Genomic profiling of CpG methylation and allelic specificity using quantitative high-throughput mass spectrometry: critical evaluation and improvements.* Nucleic Acids Res, 2007. **35**(18): p. e119.

24. Claus, R., et al., *Quantitative DNA methylation analysis identifies a single CpG dinucleotide important for ZAP-70 expression and predictive of prognosis in chronic lymphocytic leukemia.* J Clin Oncol, 2012. **30**(20): p. 2483-91.
25. Polak, J.F., et al., *Carotid-wall intima-media thickness and cardiovascular events.* N Engl J Med, 2011. **365**(3): p. 213-21.
26. Akbaraly, T.N., et al., *Adherence to healthy dietary guidelines and future depressive symptoms: evidence for sex differentials in the Whitehall II study.* Am J Clin Nutr, 2013. **97**(2): p. 419-27.
27. Agatston, A.S., et al., *Quantification of coronary artery calcium using ultrafast computed tomography.* J Am Coll Cardiol, 1990. **15**(4): p. 827-32.
28. Aryee, M.J., et al., *Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays.* Bioinformatics, 2014. **30**(10): p. 1363-9.
29. Leek, J.T., et al., *The sva package for removing batch effects and other unwanted variation in high-throughput experiments.* Bioinformatics, 2012. **28**(6): p. 882-3.
30. Steegenga, W.T., et al., *Genome-wide age-related changes in DNA methylation and gene expression in human PBMCs.* Age (Dordr), 2014. **36**(3): p. 9648.
31. Jaffe, A.E. and R.A. Irizarry, *Accounting for cellular heterogeneity is critical in epigenome-wide association studies.* Genome Biol, 2014. **15**(2): p. R31.
32. Zaina, S., et al., *DNA methylation map of human atherosclerosis.* Circ Cardiovasc Genet, 2014. **7**(5): p. 692-700.
33. Udali, S., et al., *Cardiovascular epigenetics: from DNA methylation to microRNAs.* Mol Aspects Med, 2013. **34**(4): p. 883-901.
34. Wierda, R.J., et al., *Epigenetics in atherosclerosis and inflammation.* J Cell Mol Med, 2010. **14**(6A): p. 1225-40.
35. Zaina, S., *Unraveling the DNA methylome of atherosclerosis.* Curr Opin Lipidol, 2014. **25**(2): p. 148-53.
36. Valencia-Morales Mdel, P., et al., *The DNA methylation drift of the atherosclerotic aorta increases with lesion progression.* BMC Med Genomics, 2015. **8**: p. 7.
37. Breitling, L.P., *Current genetics and epigenetics of smoking/tobacco-related cardiovascular disease.* Arterioscler Thromb Vasc Biol, 2013. **33**(7): p. 1468-72.
38. Terry, M.B., et al., *DNA methylation in white blood cells: association with risk factors in epidemiologic studies.* Epigenetics, 2011. **6**(7): p. 828-37.
39. Zhou, S., Z. Zhang, and G. Xu, *Notable epigenetic role of hyperhomocysteinemia in atherogenesis.* Lipids Health Dis, 2014. **13**: p. 134.
40. Sharma, P., et al., *Genome wide DNA methylation profiling for epigenetic alteration in coronary artery disease patients.* Gene, 2014. **541**(1): p. 31-40.
41. Guay, S.P., et al., *A study in familial hypercholesterolemia suggests reduced methylomic plasticity in men with coronary artery disease.* Epigenomics, 2015. **7**(1): p. 17-34.
42. Aavik, E., et al., *Global DNA methylation analysis of human atherosclerotic plaques reveals extensive genomic hypomethylation and reactivation at imprinted locus 14q32 involving induction of a miRNA cluster.* Eur Heart J, 2015. **36**(16): p. 993-1000.
43. Hiltunen, M.O., et al., *DNA hypomethylation and methyltransferase expression in atherosclerotic lesions.* Vasc Med, 2002. **7**(1): p. 5-11.
44. Li, X., et al., *Granulocytes in coronary thrombus evolution after myocardial infarction--time-dependent changes in expression of matrix metalloproteinases.* Cardiovasc Pathol, 2016. **25**(1): p. 40-6.
45. Spitz, C., et al., *Regulatory T cells in atherosclerosis: critical immune regulatory function and therapeutic potential.* Cell Mol Life Sci, 2016. **73**(5): p. 901-22.
46. Sharma, P., et al., *The prognostic value of BRCA1 promoter methylation in early stage triple negative breast cancer.* J Cancer Ther Res, 2014. **3**(2): p. 1-11.
47. Zhang, L. and X. Long, *Association of BRCA1 promoter methylation with sporadic breast cancers: Evidence from 40 studies.* Sci Rep, 2015. **5**: p. 17869.
48. Papoutsis, A.J., et al., *Gestational exposure to the AhR agonist 2,3,7,8-tetrachlorodibenzo-p-dioxin induces BRCA-1 promoter hypermethylation and reduces BRCA-1 expression in mammary tissue of rat offspring: preventive effects of resveratrol.* Mol Carcinog, 2015. **54**(4): p. 261-9.

49. Daniels, S.L., et al., *Levels of DNA Methylation Vary at CpG Sites across the BRCA1 Promoter, and Differ According to Triple Negative and "BRCA-Like" Status, in Both Blood and Tumour DNA.* PLoS One, 2016. **11**(7): p. e0160174.
50. Lo, P.K., et al., *Dysregulation of the BRCA1/long non-coding RNA NEAT1 signaling axis contributes to breast tumorigenesis.* Oncotarget, 2016. **7**(40): p. 65067-65089.
51. Beroud, C., et al., *BRCA Share: A Collection of Clinical BRCA Gene Variants.* Hum Mutat, 2016. **37**(12): p. 1318-1328.
52. Chehade, R., et al., *Reduced BRCA1 transcript levels in freshly isolated blood leukocytes from BRCA1 mutation carriers is mutation specific.* Breast Cancer Res, 2016. **18**(1): p. 87.
53. Diederichs, S., et al., *The dark matter of the cancer genome: aberrations in regulatory elements, untranslated regions, splice sites, non-coding RNA and synonymous mutations.* EMBO Mol Med, 2016. **8**(5): p. 442-57.
54. Wallace, A.J., *New challenges for BRCA testing: a view from the diagnostic laboratory.* Eur J Hum Genet, 2016. **24 Suppl 1**: p. S10-8.
55. Lecht, S., et al., *Anti-angiogenic activities of snake venom CRISP isolated from Echis carinatus sochureki.* Biochim Biophys Acta, 2015. **1850**(6): p. 1169-79.
56. Singh, K.K., et al., *BRCA1 is a novel target to improve endothelial dysfunction and retard atherosclerosis.* J Thorac Cardiovasc Surg, 2013. **146**(4): p. 949-960 e4.
57. Bordeleau, L., et al., *Diabetes and breast cancer among women with BRCA1 and BRCA2 mutations.* Cancer, 2011. **117**(9): p. 1812-8.
58. Moreau, K., et al., *BRCA1 affects lipid synthesis through its interaction with acetyl-CoA carboxylase.* J Biol Chem, 2006. **281**(6): p. 3172-81.
59. Shukla, P.C., et al., *BRCA1 is an essential regulator of heart function and survival following myocardial infarction.* Nat Commun, 2011. **2**: p. 593.
60. Singh, K.K., et al., *BRCA2 protein deficiency exaggerates doxorubicin-induced cardiomyocyte apoptosis and cardiac failure.* J Biol Chem, 2012. **287**(9): p. 6604-14.
61. Greenberg, R.A., *Recognition of DNA double strand breaks by the BRCA1 tumor suppressor network.* Chromosoma, 2008. **117**(4): p. 305-17.
62. Lovren, F., et al., *BRCA1 shields vascular smooth muscle cells from oxidative stress.* J Thorac Cardiovasc Surg, 2014. **147**(6): p. 1946-55, 1955 e1.
63. Jackson, K.C., et al., *BRCA1 is a novel regulator of metabolic function in skeletal muscle.* J Lipid Res, 2014. **55**(4): p. 668-80.
64. Bai, F., et al., *BRCA1 suppresses epithelial-to-mesenchymal transition and stem cell dedifferentiation during mammary and tumor development.* Cancer Res, 2014. **74**(21): p. 6161-72.
65. Chen, P.Y., et al., *Endothelial-to-mesenchymal transition drives atherosclerosis progression.* J Clin Invest, 2015. **125**(12): p. 4514-28.
66. Ma, K.L., et al., *Inflammatory stress exacerbates the progression of cardiac fibrosis in high-fat-fed apolipoprotein E knockout mice via endothelial-mesenchymal transition.* Int J Med Sci, 2013. **10**(4): p. 420-6.
67. von Gise, A. and W.T. Pu, *Endocardial and epicardial epithelial to mesenchymal transitions in heart development and disease.* Circ Res, 2012. **110**(12): p. 1628-45.
68. Liu, X., et al., *LncRNA NBR2 engages a metabolic checkpoint by regulating AMPK under energy stress.* Nat Cell Biol, 2016. **18**(4): p. 431-42.
69. Ou, H., et al., *Role of AMPK in atherosclerosis via autophagy regulation.* Sci China Life Sci, 2018. **61**(10): p. 1212-1221.
70. Ma, A., et al., *AMPK activation enhances the anti-atherogenic effects of high density lipoproteins in apoE(-/-) mice.* J Lipid Res, 2017. **58**(8): p. 1536-1547.
71. Wang, J., et al., *AMPK activation reduces the number of atheromata macrophages in ApoE deficient mice.* Atherosclerosis, 2017. **258**: p. 97-107.
72. Curjuric, I., et al., *Different genes interact with particulate matter and tobacco smoke exposure in affecting lung function decline in the general population.* PLoS One, 2012. **7**(7): p. e40175.
73. Plager, D.A., et al., *Gene transcription changes in asthmatic chronic rhinosinusitis with nasal polyps and comparison to those in atopic dermatitis.* PLoS One, 2010. **5**(7): p. e11450.

Supplementary information

Supplementary tables 1 to 6 can be found in following dropbox folder: https://www.dropbox.com/sh/d9lq9mipqkoezx3/AAC9RCXgWS7P6DS71ybJqWNJa?dl=0

Supplementary Table 1: Primer sequences.

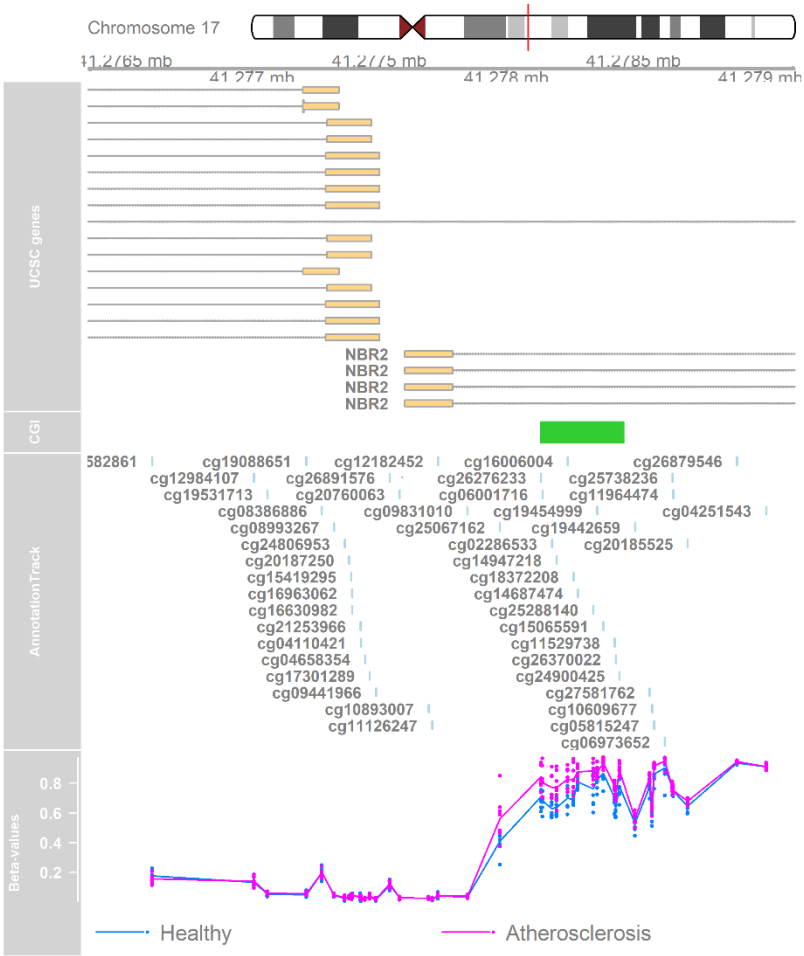
Supplementary Table 2: Significant differentially methylated positions (sig-DMPs). FDR < 0.15 and Delta_beta > 5%.

Supplementary Table 3: Significant methylated regions (sig-DMRs). Minpval < 0.001, max delta beta > 5% and ≥ 5 CpGs in region.

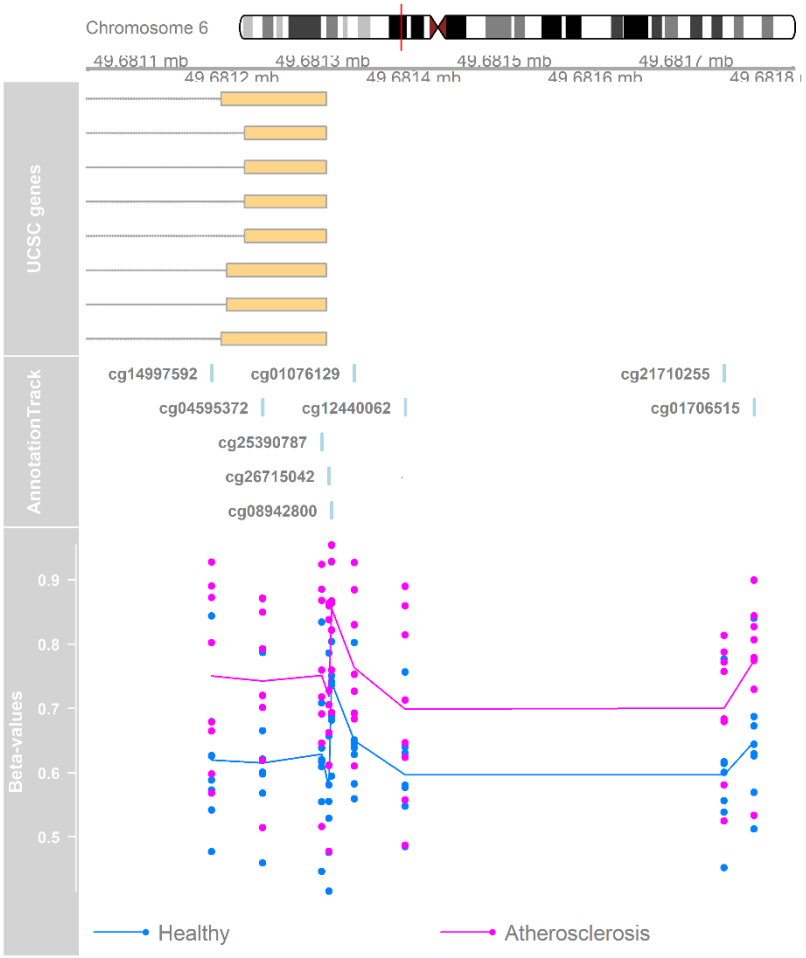
Supplementary Table 4: Relative cell type contributions in healthy and atherosclerosis group.

Supplementary Table 5: CpG probes located in sig-DMRs, associated with atherosclerosis in both blood and plaque tissues.

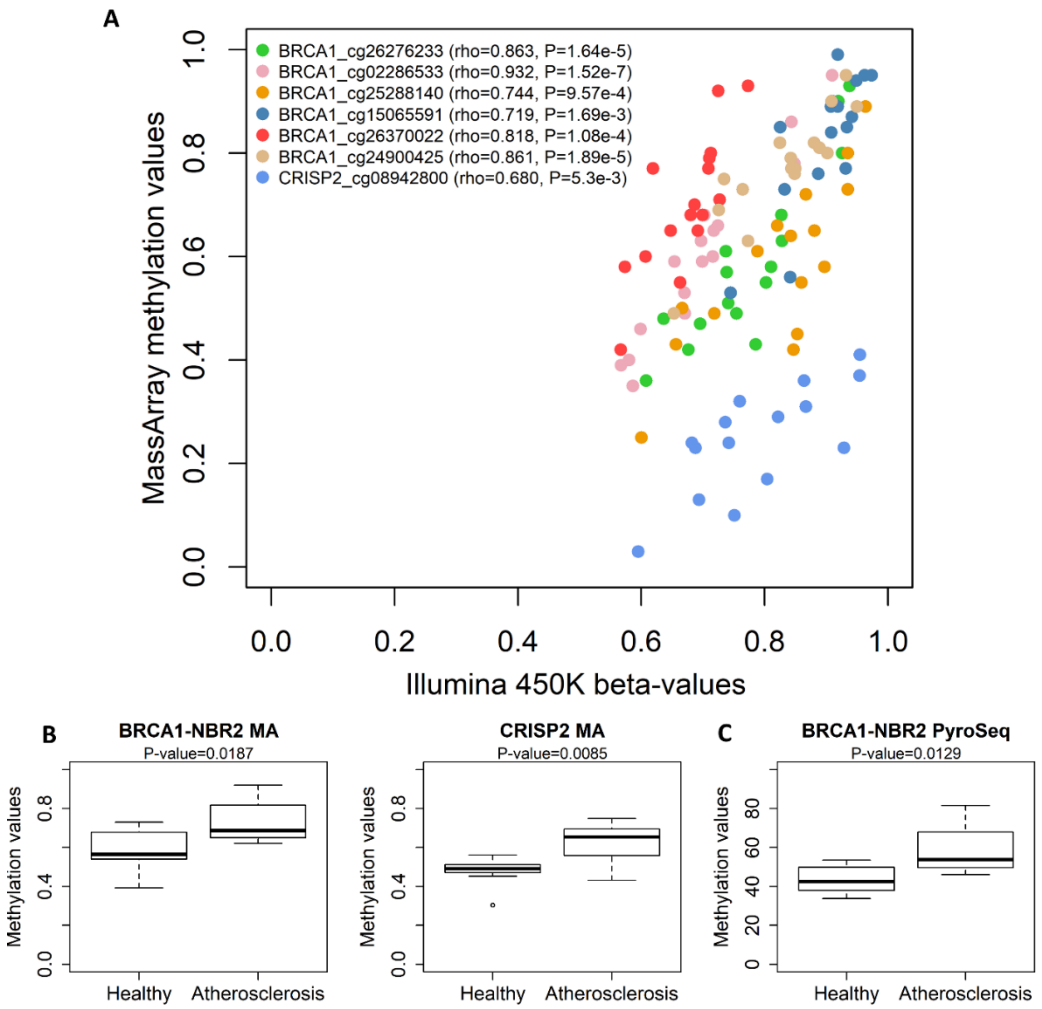
Supplementary Table 6: Ingenuity pathway enrichment analysis of gene associated DMRs of Supplementary Table 6.



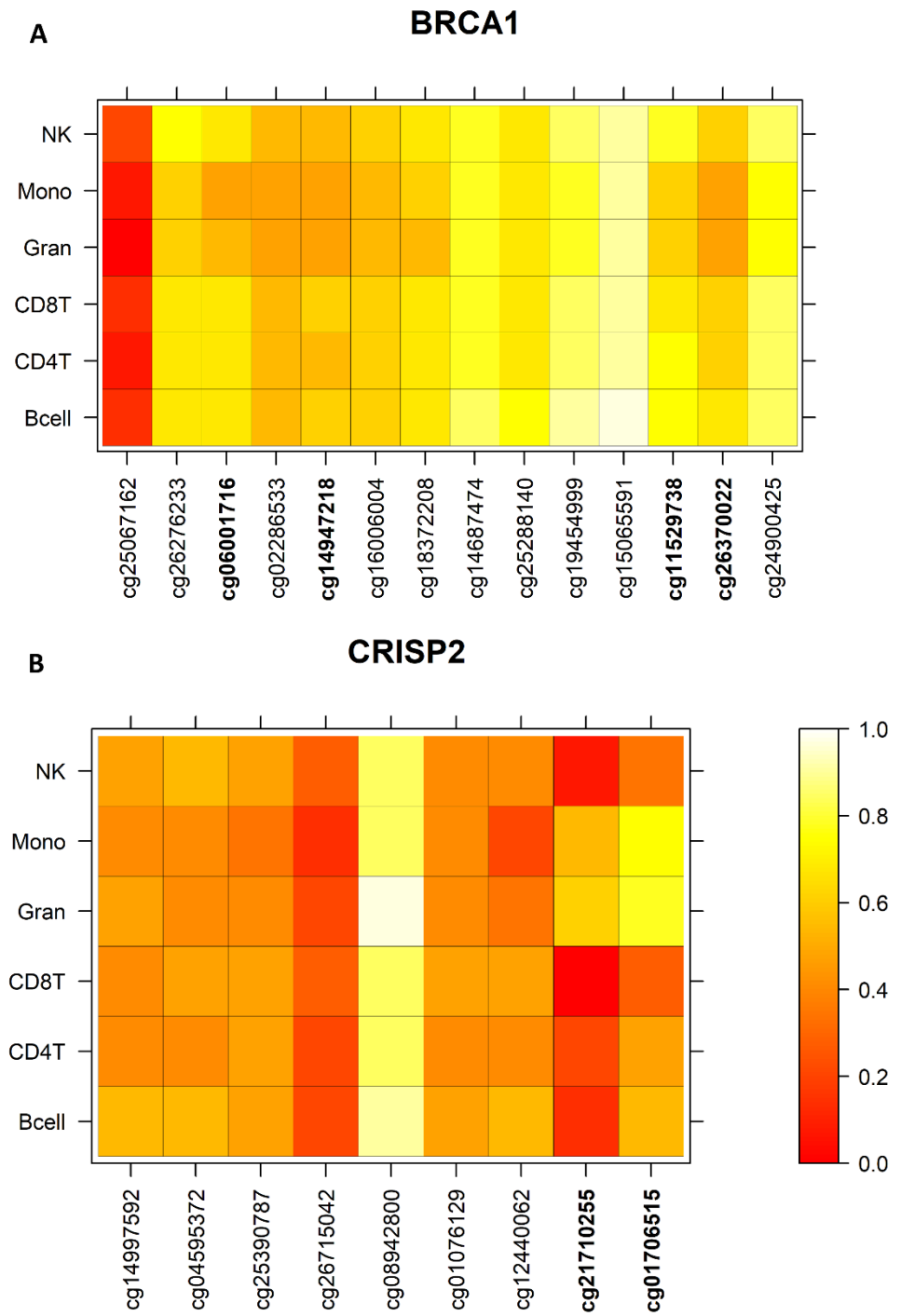
Supplementary Figure 1: UCSC genome browser view of the DMR in the promoter region of BRCA1 and NBR2.



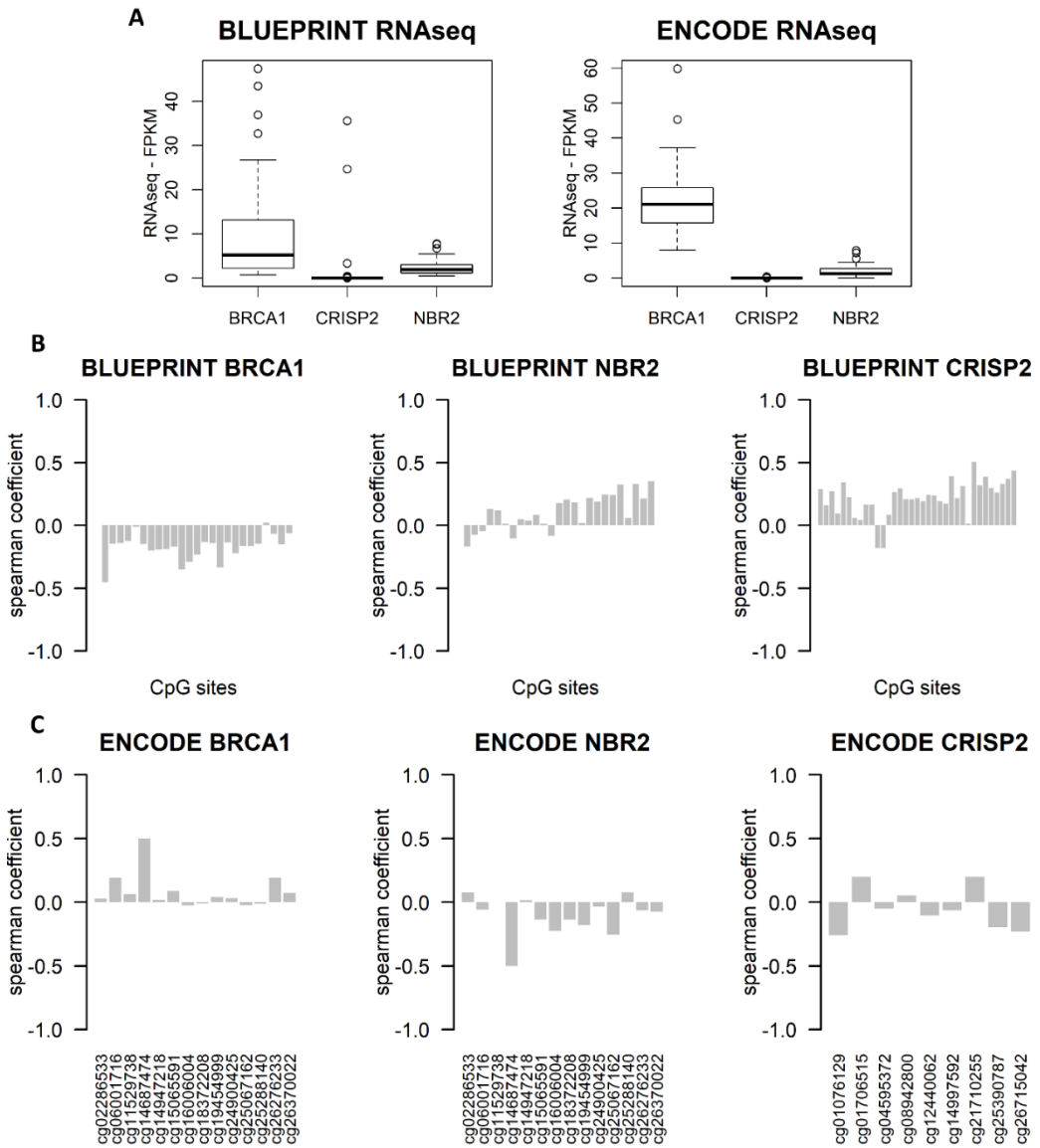
Supplementary Figure 2: UCSC genome browser view of the DMR in the promoter region of CRISP2.



Supplementary Figure 3: Verification of Illumina 450k methylation data using MassArray EpiTYPER and pyrosequencing. A) Correlation between Illumina beta-values and MassArray methylation values. Only Illumina probes are represented of which methylation was measured by a single region in the MassArray. For each gene, Spearman's correlation coefficient was calculated. **B)** MassARRAY EpiTYPER verification of BRCA1-NBR2 and CRISP2 and **C)** pyrosequencing verification of BRCA1-NBR2. The mean methylation values of each measured region are represented in boxplots. The student t-test was used to calculate the significance of the methylation difference between healthy and atherosclerotic blood samples.



Supplementary Figure 4: DNA methylation of A) BRCA1-NBR2 and B) CRISP2 DMR in different blood cell types. Color range represents DNA methylation values (beta-values). CpG probes in bold are significantly differentially methylated (p-value < 0.05) across major blood cell types determined by one-way ANOVA.



Supplementary Figure 5: Correlation between DNA methylation and gene expression using BLUEPRINT and ENCODE data. A) BRCA1, NBR2 and CRISP2 RNAseq gene expression values of BLUEPRINT (left) and ENCODE (right) samples. **B)** Correlations between DNA methylation and gene expression in BLUEPRINT samples. **C)** Correlations between DNA methylation and gene expression in ENCODE samples.

6

Characterization of blood surrogate immune-methylation biomarkers for immune cell infiltration in chronic inflammaging disorders

Ken Declerck[1] and Wim Vanden Berghe[1]

[1] Laboratory of protein chemistry, Proteomics and Epigenetic Signaling (PPES), Department of Biomedical Sciences, Faculty of Pharmaceutical, Biomedical and Veterinary Sciences, Antwerp University, Antwerp (Wilrijk), Belgium

Under revision in Frontiers in Genetics.

Contributions: Ken Declerck analyzed the data, interpreted the results and wrote the manuscript.

Abstract | Alzheimer's disease (AD) and atherosclerosis are both chronic age- and inflammation-dependent diseases. In addition, atherosclerosis is frequently observed in AD patients indicating common involvement of vascular components in disease etiologies. Recently, epigenome-wide association studies have identified epigenetic alterations, and in particularly DNA methylation changes for both diseases. We hypothesized the existence of a common DNA methylation profile in atherosclerosis and AD which may be valuable as a blood-based DNA methylation inflammaging biomarker.

Using publicly available 450k Illumina methylation datasets, we identified a co-methylation network associated with both atherosclerosis and AD in whole blood samples. This methylation profile appeared to indicate shifts in blood immune cell type distribution. Remarkably, similar methylation changes were also detected in disease tissues, including AD brain tissues, atherosclerotic plaques and tumors and were found to correlate with immune cell infiltration. In addition, this immune-related methylation profile could also be detected in other inflammaging diseases, including Parkinson's disease and obesity, but not in multiple sclerosis, schizophrenia and osteoporosis.

In conclusion, we identified a blood-based immune-related DNA methylation signature in multiple inflammaging diseases associated with changes in blood immune cell counts and predictive for immune cell infiltration in diseased tissues. In addition to epigenetic clock measurements, this immune-methylation signature may become a valuable blood-based biomarker to prevent chronic inflammatory disease development or monitor lifestyle intervention strategies which promote healthy aging.

Introduction

Aging and inflammation are important contributors of various chronic lifestyle diseases, including Alzheimer's disease (AD) and atherosclerosis. Furthermore, AD and atherosclerosis share a lot of disease characteristics and it has been hypothesized that they have a common cause [1].

AD is the most common form of dementia, and is characterized by the accumulation and aggregation of extracellular amyloid-β (Aβ) plaques, the intraneuronal deposition of hyper-phosphorylated tau protein which forms neurofibrillary tangles, neuronal loss and gliosis in the cerebral cortex and hippocampus [2, 3]. In addition, also vascular components seem to play a crucial role in the initiation and development of AD [4-6]. The brain consumes a high amount of oxygen and glucose, and therefore the cerebral blood flow is of particular importance for brain health. It is therefore not surprising that cerebrovascular dysfunction has been associated with dementia, AD and other

neurodegenerative disorders. More recently, it has been hypothesized that cerebrovascular damage could be the first hit in AD initiation leading to neuronal injury and loss, and the accumulation of Aβ in the brain, and eventually the development of AD [6].

Atherosclerosis has been associated with dementia and AD [1, 7]. During atherosclerosis development, lipids, macrophages, fibrous connective tissue and necrotic debris accumulate in the arteries wall leading to the formation of plaques, which can over time rupture and block the blood flow leading eventually to CVDs like myocardial infarction or stroke [8]. Both atherosclerosis as AD risk increases with age and have an inflammatory component. Of interest, cerebrovascular atherosclerosis has been found to occur more often in AD patients and correlate with the severity of cognitive impairment [9-17]. Also carotid atherosclerosis, carotid intima media thickness and coronary artery disease has been associated with AD and AD pathology [18-22]. In addition, adults with CVD show an increased risk for the development of dementia and AD [23]. Furthermore, atherosclerosis and AD share common risk factors including age, hypertension, type 2 diabetes, obesity, smoking, hypercholesterolemia and hyperhomocysteinemia [2, 24, 25]. Of interest, in both diseases the APOE4 allele is a genetic risk factor [26, 27].

Because both diseases are associated with multiple lifestyle and environmental factors, it is not surprising that epigenetic mechanisms are involved in both disease etiologies. Epigenetics is linking environmental factors and genetics through modulation of gene expression patterns. Blood and saliva DNA methylation profiles are increasingly applied as valuable diagnostic and prognostic biomarkers in diseased patients. DNA methylation alterations have been identified in whole blood and plaque tissues of atherosclerosis [28-34]. Also AD has been associated with methylation changes in blood and different brain regions [35, 36]. We recently demonstrated that BRCA1 and CRISPR specific DNA changes in blood can be used as surrogate marker for atherosclerosis [28]. More particularly, hypermethylation of a CGI in the promoter region of BRCA1 could be replicated in plaque tissue of two independent cohorts indicating that blood can be used to predict methylation changes in atherosclerotic plaques. Of interest, BRCA1 promoter was also found to be differentially methylated in AD within neurons, and found to be correlated with gene expression [37]. In AD, however, there is limited evidence that methylation changes in brain tissues are also present in more accessible tissues like blood [38, 39]. In a study of Lunnon and colleagues, methylation changes found in blood of AD patients were not overlapping with the changes seen in AD brain [36]. However, the AD blood DMPs were located in the vicinity of genes of relevance to AD and correlated with transcriptional changes making them still potential diagnostic biomarkers.

Given the high commonalities between atherosclerosis and AD disease , here we further examined whether we could find similar DNA methylation signatures in blood of AD and atherosclerosis patients.

Materials and methods

Datasets

Genome-wide DNA methylation datasets were extracted from the GEO database. The datasets contain genome-wide DNA methylation profiles of AD brain and whole blood samples, and plaques and whole blood samples of atherosclerotic patients. Also, one dataset containing samples of intracranial aneurysm arteries was included. Genome-wide methylation levels were measured using the 450k Illumina arrays in every dataset. Table 1 summarizes all AD related datasets used in our study. The dataset_ID is used to refer to each dataset in the main text. Table 2 summarizes chronic disease related datasets (Table 2).

Table 1: GEO methylation datasets of AD.

Dataset_ID	GEO accession	Disease	Tissue
AD_cerebellum_GSE59685	GSE59685	AD	Cerebellum
AD_EntorhinalCortex_GSE59685	GSE59685	AD	Entorhinal cortex
AD_FrontalCortex_GSE59685	GSE59685	AD	Frontal cortex
AD_SupTempGyrus_GSE59685	GSE59685	AD	Superior temporal gyrus
AD_wholeblood_GSE59685	GSE59685	AD	Whole blood
AD_cerebellum_GSE72778	GSE72778	AD	Cerebellum
AD_Frontal_GSE72778	GSE72778	AD	Frontal cortex
AD_Hippocampus_GSE72778	GSE72778	AD	Hippocampus
AD_Occipital_GSE72778	GSE72778	AD	Occipital cortex
AD_TemporalCortex_GSE72778	GSE72778	AD	Temporal cortex
AD_PrefrontalCortex_GSE80970	GSE80970	AD	Prefrontal cortex
AD_SupTempGyrus_GSE80970	GSE80970	AD	Superior temporal gyrus
AD_SupTempGyrus_GSE76105	GSE76105	AD	Superior temporal gyrus
Athero_wholeblood_GSE107143	GSE107143	Atherosclerosis	Whole blood
IntracranAneurysm_artery_GSE75434	GSE75434	Intracranial aneurysm	Superficial temporal artery
AtheroCerebrovas_plaque_GSE66500	GSE66500	Atherosclerosis with cerebrovascular event	Carotid plaque

Athero_plaquePaired_GSE46394	GSE46394	Atherosclerosis	Aortic plaque
Athero_plaque_GSE46394	GSE46394	Atherosclerosis	Carotid plaque

Table 2: GEO whole blood methylation datasets of different inflammaging diseases.

Accession_ID	Platform	Tissue	Disease	Abbreviation
GSE107143	450k	Whole Blood	Atherosclerosis	athero
GSE59685	450k	Whole Blood	Alzheimer's Disease	AD
GSE72774	450k	Whole Blood	Parkinson's Disease	PD
GSE88824	450k	Whole Blood	Multiple Sclerosis	MS
GSE73103	450k	Whole Blood	Obesity	obese
GSE41169	450k	Whole Blood	Schizophrenia	schizo
GSE99624	450k	Whole Blood	Osteoporosis	osteo

Weighted correlation network analysis (WGCNA)

To detect consensus modules between atherosclerosis and AD in whole blood, the WGCNA R package was used. First, the most variable probes were selected based on an median absolute deviation (MAD) threshold of 0.03 in at least one dataset. In this way 97,375 probes remained for further analysis. The blockwiseConsensusModules function in the WGCNA package was subsequently used to detect consensus modules across the two datasets. We used the soft-threshold power of 7, a minimum module size of 30 probes, a maximum block size of 20,000 and a dendrogram cut height of 0.25 for module merging as input parameters. The consensus module eigengenes was associated with disease (either atherosclerosis or AD), and the modules with a significant association (p-value < 0.05) in both datasets were used for further analysis. The module membership of each probe in the modules was calculated by correlating the module eigengenes with the beta values. The gene significance values of each probe in the modules were calculated using the -log10(p-value) of the association between the beta-values and the disease groups. Probes in the significant modules were mapped to different genomic regions, including gene elements, CGI elements and chromatin segmentation states. The enrichment of module probes in one of the genomic regions was calculated using the fisher's Exact test. Probes in the significant modules were mapped to genes using the Illumina manifest annotation file. Pathway enrichment was performed using the IPA software. Module preservation across different AD and cardiovascular datasets were performed using the modulePreservation function in the WGCNA R package. 100 permutations were performed to calculate the preservation z-scores for each dataset. Z-scores higher than 10 indicate strong preservation, between 2 and 10 weak to moderate preservation and below 2 no preservation.

Estimation of cell counts and immune cell infiltration

Cell type fractions were calculated using the method described by Houseman et al. The EpiDISH R package was used to perform the calculations. For the whole blood datasets, we used the centDHSbloodDMC.m whole blood reference dataset containing 333 CpG probes. To estimate the cell counts in the atherosclerosis vascular tissues, we created a new reference methylome. For the smooth muscle cells, fibroblasts and endothelial cells, we retrieved 450k methylation data of aortic smooth muscle cells (AoSMC), ProgFib and human umbilical vein endothelial cells (HUVEC) from the ENCODE project (GSE40699), respectively. Immune cell reference methylomes were obtained from the study of Reinius et al. Next, differences in methylation across the different cell types were calculated using limma linear models comparing each cell type with the rest of the samples: immune cell vs rest of samples, AoSMC vs rest of samples, ProfFib vs rest of samples and HUVEC vs rest of samples. For each cell type the top 100 significant CpG probes with the largest methylation difference were selected and combined to obtain 357 unique CpG probes. The beta values of the ICs were averaged to obtain the final reference methylome. This reference methylome was subsequently used to estimate cell counts and IC infiltration in the vascular tissues. Information about IC infiltration of the TCGA cancers were obtained from a recent study examining immunogenomic profiles of different cancers [40]. TCGA 450k Illumina methylation data were retrieved using the TCGAbiolinks R package.

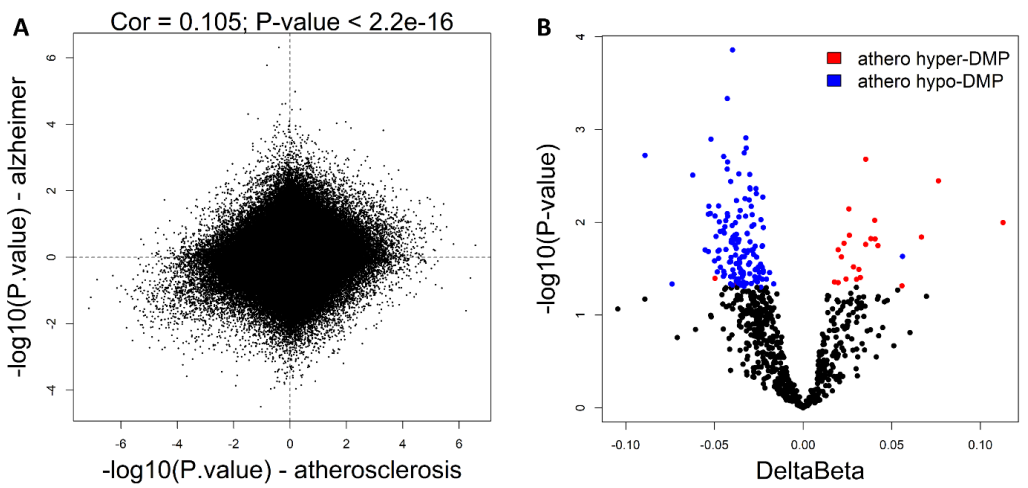
Results

Atherosclerosis and AD whole blood samples contain a common DNA methylation signature

To compare methylation profiles between atherosclerosis and AD in whole blood, we first compared the genome-wide significance of each CpG probe in both datasets. Using the limma moderated t-test, we performed differentially methylation analysis on both whole blood datasets. Next, the -log10(p-value) for each Illumina CpG-probe was calculated, and was made negative for hypomethylated CpG probes. We found a weak positive correlation between the -log10(p-values) in both datasets, indicating that at the genome-wide level the similarity between the methylation profiles in atherosclerosis and AD is limited (**Figure 1A**). Next, we checked more specifically, whether the top significant CpG-probes found in our atherosclerosis dataset were also differentially methylated in the AD dataset. We first selected the most significant CpG-

probes by setting the threshold for differentially methylation at FDR < 0.15 and delta beta > 0.05. In this way 712 CpG-probes were selected which we called athero-DMPs. T-tests were performed to determine the significance level in the AD dataset for each of the 712 athero-DMPs. We found several probes which were also found to be differentially methylated in AD (**Figure 1B**). Of particular interest, the directionality of the methylation change was very similar in both datasets.

Using different publicly available DNA methylation datasets, we performed the same analyses in multiple AD brain tissues. Again, for some of the brain tissue we could find a similar methylation profile compared to atherosclerosis blood samples (**Figure 1C-D and Supplementary Figure 1**). Especially in frontal lobe, frontal cortex and superior temporal gyrus, the hypo- and hypermethylated atherosclerosis DMPs corresponded with hypo- and hypermethylation in AD, respectively (**Supplementary Figure 1**). Except for cerebellum tissues, there was a positive correlation between the -log10(p-values) of the athero-DMPs in the atherosclerosis whole blood dataset and the other AD brain and whole blood datasets (**Figure 1C**). In contrast, in cerebellum samples no strong correlation could be found. Furthermore, cerebellum samples did not correlate with the other brain AD tissues (**Figure 1D**).



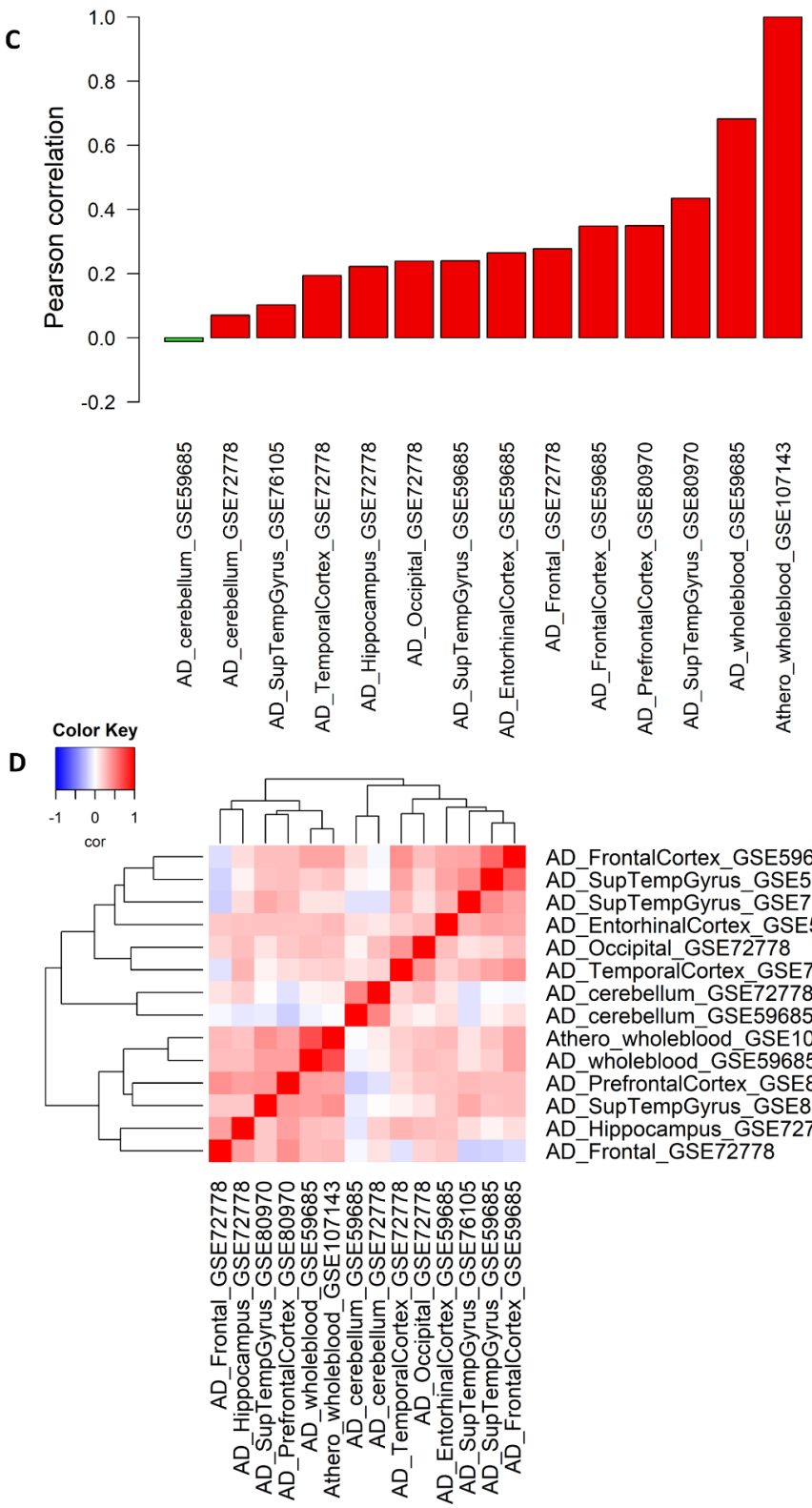

C
Pearson correlation
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
AD_cerebellum_GSE59685
AD_cerebellum_GSE72778
AD_SupTempGyrus_GSE76105
AD_TemporalCortex_GSE72778
AD_Hippocampus_GSE72778
AD_Occipital_GSE72778
AD_SupTempGyrus_GSE59685
AD_EntorhinalCortex_GSE59685
AD_Frontal_GSE72778
AD_FrontalCortex_GSE59685
AD_PrefrontalCortex_GSE80970
AD_SupTempGyrus_GSE80970
AD_wholeblood_GSE59685
Athero_wholeblood_GSE107143
D
Color Key
-1
0
1
cor
AD_FrontalCortex_GSE59685
AD_SupTempGyrus_GSE59685
AD_SupTempGyrus_GSE76105
AD_EntorhinalCortex_GSE59685
AD_Occipital_GSE72778
AD_TemporalCortex_GSE72778
AD_cerebellum_GSE72778
AD_cerebellum_GSE59685
Athero_wholeblood_GSE107143
AD_wholeblood_GSE59685
AD_PrefrontalCortex_GSE80970
AD_SupTempGyrus_GSE80970
AD_Hippocampus_GSE72778
AD_Frontal_GSE72778
AD_Frontal_GSE72778
AD_Hippocampus_GSE72778
AD_SupTempGyrus_GSE80970
AD_PrefrontalCortex_GSE80970
AD_wholeblood_GSE59685
Athero_wholeblood_GSE107143
AD_cerebellum_GSE59685
AD_cerebellum_GSE72778
AD_TemporalCortex_GSE72778
AD_Occipital_GSE72778
AD_EntorhinalCortex_GSE59685
AD_SupTempGyrus_GSE76105
AD_SupTempGyrus_GSE59685
AD_FrontalCortex_GSE59685

Figure 1: DNA methylation similarity in atherosclerosis and AD whole blood and brain samples. A) The genome-wide significance levels (-log10(P.value)) of each CpG-probe in the atherosclerosis and AD whole blood dataset was plotted in the X-axis and Y-axis, respectively. The significance levels were made negative for hypomethylated probes. The Pearson correlation was used to calculate the correlation between the two datasets. **B)** Volcanoplot showing the methylation differences of the 712 top significant atherosclerosis DMPs in the AD whole blood dataset (athero-DMPs). Significant probes (p-value < 0.05) were colored blue when hypomethylated and red when hypermethylated in atherosclerosis. **C)** Correlation coefficients between -log10(p-values) of athero-DMPs in the atherosclerosis whole blood dataset and multiple AD brain and whole blood datasets. **D)** Correlation heatmap of the -log10(p-values) for the athero-DMPs across different AD datasets. Red means a positive correlation and blue a negative correlation.

Next, we calculated the gene significance and module membership of the probes in the consensus modules. We defined the gene significance as the -log10(p-value) of the association between the CpG-probe methylation value and disease state, and the module membership as the correlation coefficient between the module eigengene and the CpG-probe methylation value. The closer the module membership is to 1 or -1 the more important the probe is in the module. In general, a module membership close to 1 or -1 is highly connective and therefore represents a hub in the network. There was a strong correlation between the gene significance in the two datasets for module ME21 (Pearson's correlation: 0.845) (**Figure 2A**). The same was true for the module membership (Pearson's correlation: 0.989) (**Figure 2B**). As expected the gene significance and module membership was also highly correlated (**Supplementary Figure 2**). Because a less strong correlation could be found with module ME54 (data not shown), we decided to focus only on module ME21.

The atherosclerosis-AD blood consensus network is also associated in brain tissues and atherosclerotic plaques

We further analyzed whether module 21 was preserved in other AD methylation datasets of different brain tissues. The preservation z-scores for all AD brain tissues, except for cerebellum, were between 2 and 10, suggesting weak to moderate preservation (**Figure 3A**). In cerebellum, there was no indication of module preservation (z-score < 2). We next calculated for each AD dataset the gene significance values of the CpG probes in module 21, and performed pairwise correlation across the different AD datasets. Except for cerebellum, all the other AD datasets showed a positive correlation with the gene significance values of the whole blood datasets, and relative to each other (**Figure 3B**).

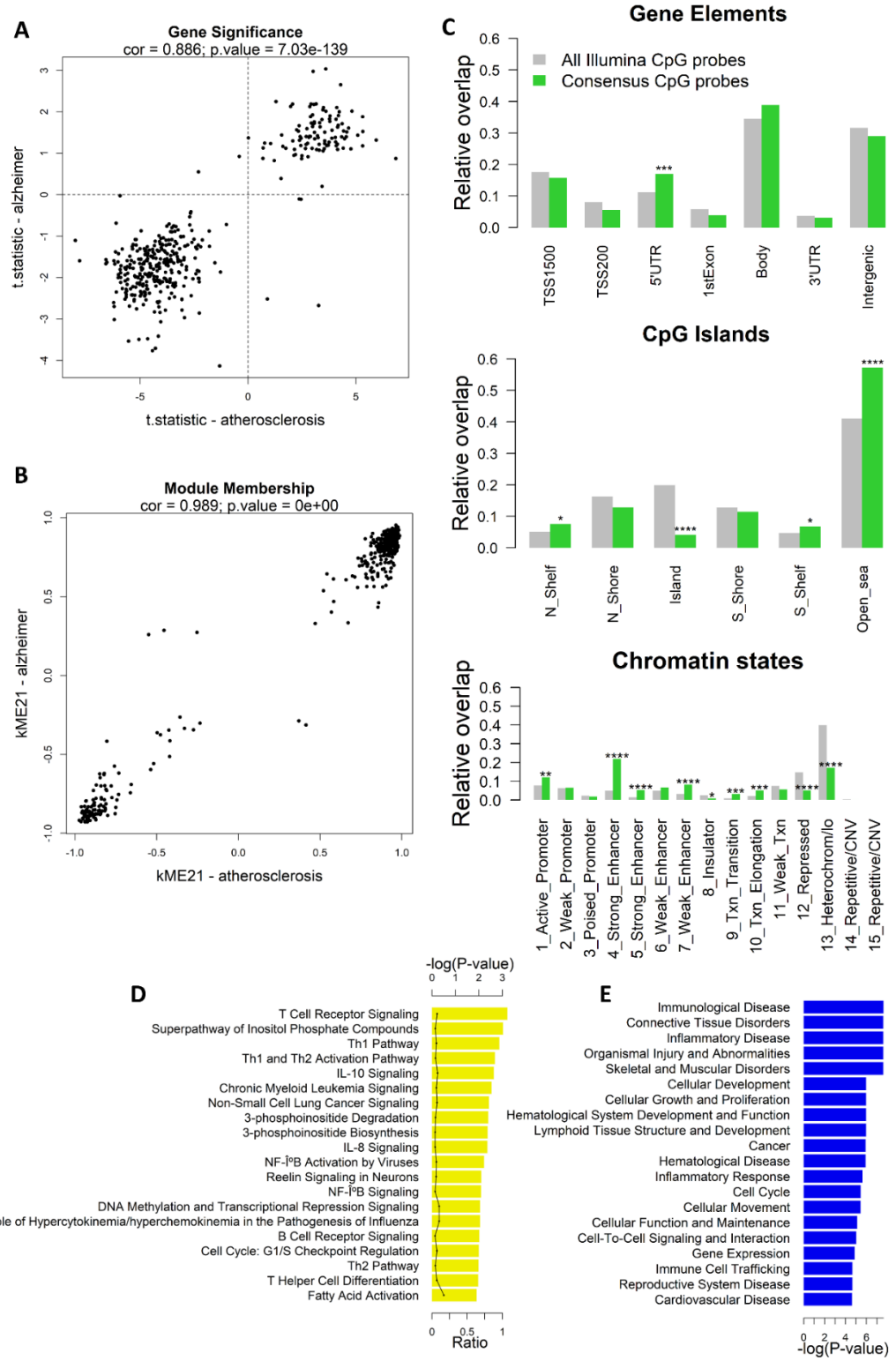
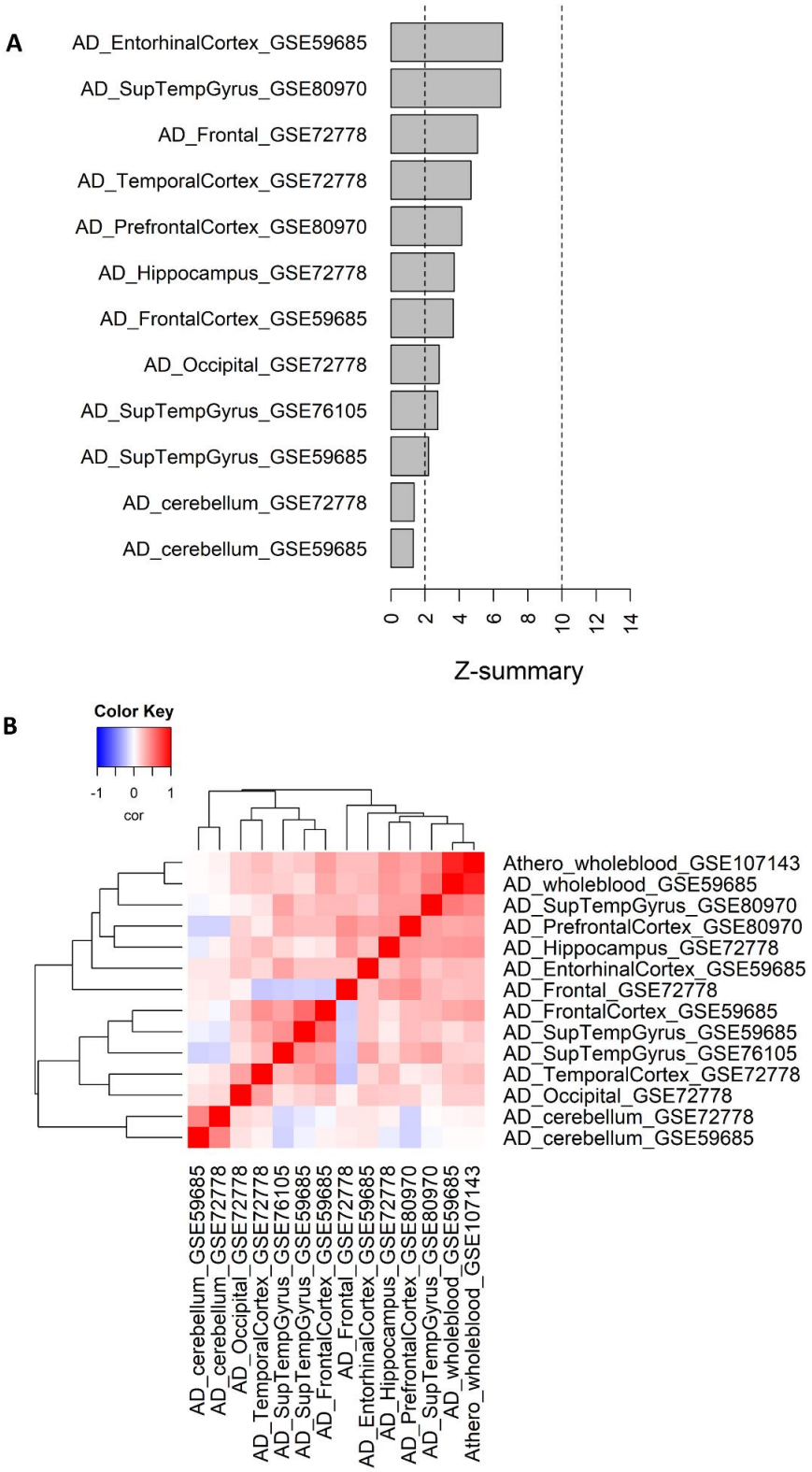

A
Gene Significance
cor = 0.886; p.value = 7.03e-139
t.statistic - alzheimer
t.statistic - atherosclerosis
B
Module Membership
cor = 0.989; p.value = 0e+00
kME21 - alzheimer
kME21 - atherosclerosis
C
Gene Elements
All Illumina CpG probes
Consensus CpG probes
Relative overlap
TSS1500
TSS200
5'UTR
1stExon
Body
3'UTR
Intergenic
CpG Islands
N_Shelf
N_Shore
Island
S_Shore
S_Shelf
Open_sea
Chromatin states
1_Active_Promoter
2_Weak_Promoter
3_Poised_Promoter
4_Strong_Enhancer
5_Strong_Enhancer
6_Weak_Enhancer
7_Weak_Enhancer
8_Insulator
9_Txn_Transition
10_Txn_Elongation
11_Weak_Txn
12_Repressed
13_Heterochrom/lo
14_Repetitive/CNV
15_Repetitive/CNV
D
-log(P-value)
T Cell Receptor Signaling
Superpathway of Inositol Phosphate Compounds
Th1 Pathway
Th1 and Th2 Activation Pathway
IL-10 Signaling
Chronic Myeloid Leukemia Signaling
Non-Small Cell Lung Cancer Signaling
3-phosphoinositide Degradation
3-phosphoinositide Biosynthesis
IL-8 Signaling
NF-ÎºB Activation by Viruses
Reelin Signaling in Neurons
NF-ÎºB Signaling
DNA Methylation and Transcriptional Repression Signaling
Role of Hypercytokinemia/hyperchemokinemia in the Pathogenesis of Influenza
B Cell Receptor Signaling
Cell Cycle: G1/S Checkpoint Regulation
Th2 Pathway
T Helper Cell Differentiation
Fatty Acid Activation
Ratio
E
Immunological Disease
Connective Tissue Disorders
Inflammatory Disease
Organismal Injury and Abnormalities
Skeletal and Muscular Disorders
Cellular Development
Cellular Growth and Proliferation
Hematological System Development and Function
Lymphoid Tissue Structure and Development
Cancer
Hematological Disease
Inflammatory Response
Cell Cycle
Cellular Movement
Cellular Function and Maintenance
Cell-To-Cell Signaling and Interaction
Gene Expression
Immune Cell Trafficking
Reproductive System Disease
Cardiovascular Disease
-log(P-value)

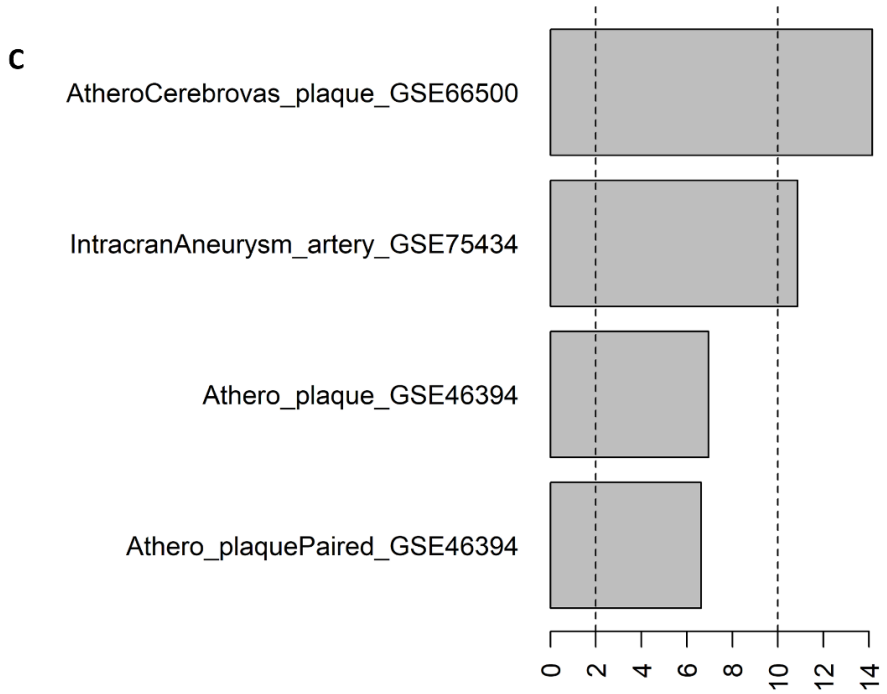
Figure 2: WGCNA co-methylation consensus modules in atherosclerosis and AD whole blood datasets. **A)** Correlation of gene significance values (t-statistics) of CpG probes in the consensus module between atherosclerosis and AD whole blood datasets. Heatmap showing the consensus modules of which the module eigengenes were significantly associated with either atherosclerosis or AD. Significant modules were marked with an asterisk. Red means a positive association and blue a negative. **B)** Correlation of gene significance (left) and module membership (right) of CpG probes in the consensus module 21 between atherosclerosis and AD whole blood datasets. **C)** Genomic enrichment of the consensus module CpG probes in multiple genomic regions: gene elements (top), CGI elements (center) and chromatin segmentation states (bottom). * Fisher's Exact $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, **** $P \leq 0.0001$. module 21 CpG probes. Barplots representing the log2 enrichment ratios of the module 21 CpG probes in multiple genomic regions: gene elements (top left), CGIs (top right), chromatin segmentation states (bottom left) and TFBS (bottom right). **D)** Significantly enriched IPA canonical pathways, and (E) IPA diseases and biofunctions of genes containing a consensus module CpG probe.

We next wondered whether the same pattern could also be found in other methylation datasets related to CVD and atherosclerosis. We extracted 450k Illumina data from carotid plaques, plaques after cerebrovascular event and arteries with intracranial aneurysm. Here the preservation was much stronger, with z-scores higher than 10 in the atherosclerotic plaques after a cerebrovascular event and in intracranial aneurysm, while for the carotid plaque datasets we found moderated module preservation (**Figure 3C**). Again, we could find strong positive correlations between the gene significance values of the different datasets (**Figure 3D**). Of note, the highest correlation with the whole blood dataset could be found with the atherosclerotic plaque dataset with a cerebrovascular event. In contrast, there was no evidence of correlation between the carotid plaque dataset and the dataset with a cerebrovascular event.

We next mapped the CpG probes in module 21 to different genomic regions relative to gene elements (TSS, gene bodies, …), CGIs, and chromatin segmentation states. Interestingly, we found an enrichment in 5'UTR regions, CpG-poor regions outside CGIs, active promoters, strong and weak enhancers, transcriptional transition and elongation states (**Figure 2C**). In addition, module 21 probes were strongly depleted in CGIs, repressed chromatin states and heterochromatin.

The CpG probes in module 21 were subsequently mapped to genes. IPA pathway analysis showed a strong enrichment in T cell regulatory and immune pathways, including T- and B cell receptor signaling, Th1 and Th2 pathway, IL-10 and IL-8 signaling, and NF-κB signaling (**Figure 2D**). In addition, top enriched diseases and biofunction were immunological and inflammatory diseases, and functions related to cellular development, growth, proliferation and movement (**Figure 2E**).


A
AD_EntorhinalCortex_GSE59685
AD_SupTempGyrus_GSE80970
AD_Frontal_GSE72778
AD_TemporalCortex_GSE72778
AD_PrefrontalCortex_GSE80970
AD_Hippocampus_GSE72778
AD_FrontalCortex_GSE59685
AD_Occipital_GSE72778
AD_SupTempGyrus_GSE76105
AD_SupTempGyrus_GSE59685
AD_cerebellum_GSE72778
AD_cerebellum_GSE59685
0
2
4
6
8
10
12
14
Z-summary
B
Color Key
-1
0
1
cor
Athero_wholeblood_GSE107143
AD_wholeblood_GSE59685
AD_SupTempGyrus_GSE80970
AD_PrefrontalCortex_GSE80970
AD_Hippocampus_GSE72778
AD_EntorhinalCortex_GSE59685
AD_Frontal_GSE72778
AD_FrontalCortex_GSE59685
AD_SupTempGyrus_GSE59685
AD_SupTempGyrus_GSE76105
AD_TemporalCortex_GSE72778
AD_Occipital_GSE72778
AD_cerebellum_GSE72778
AD_cerebellum_GSE59685
AD_cerebellum_GSE59685
AD_cerebellum_GSE72778
AD_Occipital_GSE72778
AD_TemporalCortex_GSE72778
AD_SupTempGyrus_GSE76105
AD_SupTempGyrus_GSE59685
AD_FrontalCortex_GSE59685
AD_Frontal_GSE72778
AD_EntorhinalCortex_GSE59685
AD_Hippocampus_GSE72778
AD_PrefrontalCortex_GSE80970
AD_SupTempGyrus_GSE80970
AD_wholeblood_GSE59685
Athero_wholeblood_GSE107143


C
AtheroCerebrovas_plaque_GSE66500
IntracranAneurysm_artery_GSE75434
Athero_plaque_GSE46394
Athero_plaquePaired_GSE46394
0
2
4
6
8
10
12
14
Z-summary

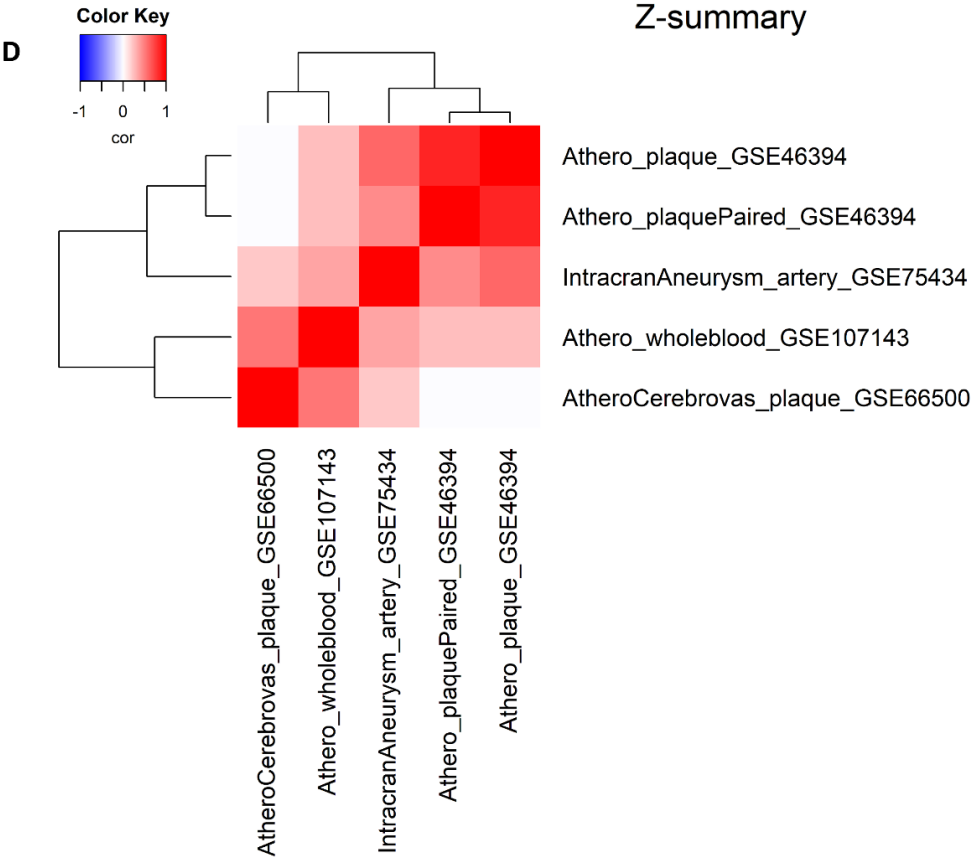

D
Color Key
-1
0
1
cor
Athero_plaque_GSE46394
Athero_plaquePaired_GSE46394
IntracranAneurysm_artery_GSE75434
Athero_wholeblood_GSE107143
AtheroCerebrovas_plaque_GSE66500
AtheroCerebrovas_plaque_GSE66500
Athero_wholeblood_GSE107143
IntracranAneurysm_artery_GSE75434
Athero_plaquePaired_GSE46394
Athero_plaque_GSE46394

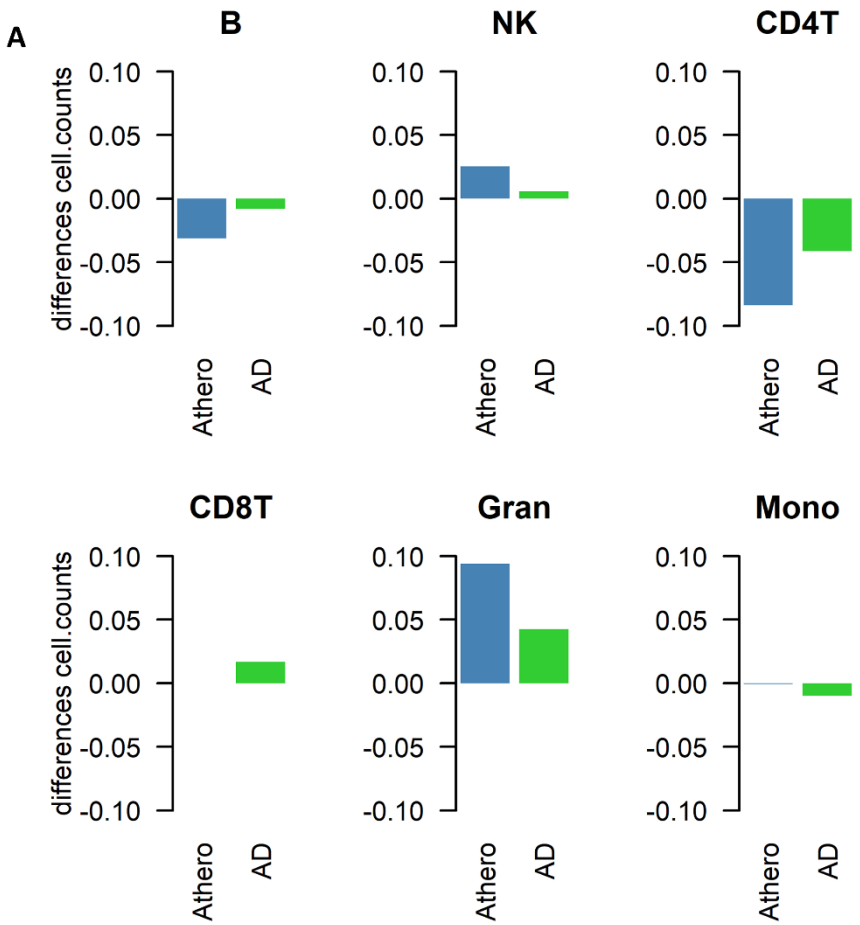
Figure 3: WGCNA consensus module in AD brain tissues and atherosclerotic plaques. A) Module preservation in AD tissue datasets. For each dataset the preservation z-score is shown. Z-scores below indicates no preservation, between two and ten weak to moderate preservation and above ten strong preservation. **B)** Correlation heatmap of -log10(p-values) significance levels between different AD datasets. **C)** Module preservation in atherosclerosis and CVD datasets. **D)** Correlation heatmap of -log10(p-values) significance levels between different CVD datasets.

The atherosclerosis-AD blood consensus network represents a common immuno-methylation signature

The enriched pathways in T cell activation and function indicate that part of the methylation changes may be due to differences in cell type heterogeneity in the samples analyzed. We therefore estimated cell type composition in the atherosclerosis and AD whole blood datasets. B-cells and CD4 T-cell levels were both reduced while granulocyte levels were increased in atherosclerosis and AD samples as compared to healthy blood samples (**Figure 4A**). In addition, many CpG probes in the consensus module seem variable across the different blood immune cell types (**Supplementary Figure 3**). In this respect, the positive correlations seen with the AD brain and atherosclerotic plaque tissues could reflect the infiltration of immune cells in the brain and arterial wall respectively. To verify this hypothesis, we estimated immune cell fraction in the vascular tissues using a new reference methylome created from methylation profiles of AoSMC, fibroblasts (ProgFib), endothelials cells (HUVEC) and immune cells (IC) (see method section for detials). As expected, an increase in IC was observed in plaque tissue compared to healthy aorta tissue (**Figure 4B**), and monocytes were the main infiltrated blood cell type observed in the plaques (**Supplementary Figure 4**). More surprisingly, the AoSMC fraction was relatively lowered in plaque material. HUVEC and ProgFib fractions didn't show substantial differences. Methylation values of the 500 most significantly differentially methylated probes in aorta plaques from the reference methylomes revealed that the hypermethylated profile was mainly due to an overall hypermethylation in immune cells compared to the other cell types. Similarly, the small fraction of hypo-DMPs could als be attributed to hypomethylated CpG sites in immune cells (**Supplementary Figure 5**). Furthermore, a strong correlation was found between the consensus module eigengene and the estimated immune cell fraction, supporting our hypothesis (**Figure 4C**).

To further prove that our methylation profile measures an immune component, we made use of immune cell infiltration information of TCGA cancers obtained from a recent study [40]. As expected, in almost all cancers there was a negative correlation between the module eigengenes and leukocyte fraction, stromal fraction and lymphocyte infiltration signature score (**Supplementary Figure 6**). Thus, tumors with methylation profiles resembling the methylation consensus module demonstrated

more immune cell infiltration. This is completely in line with our observations and supports our conclusions.



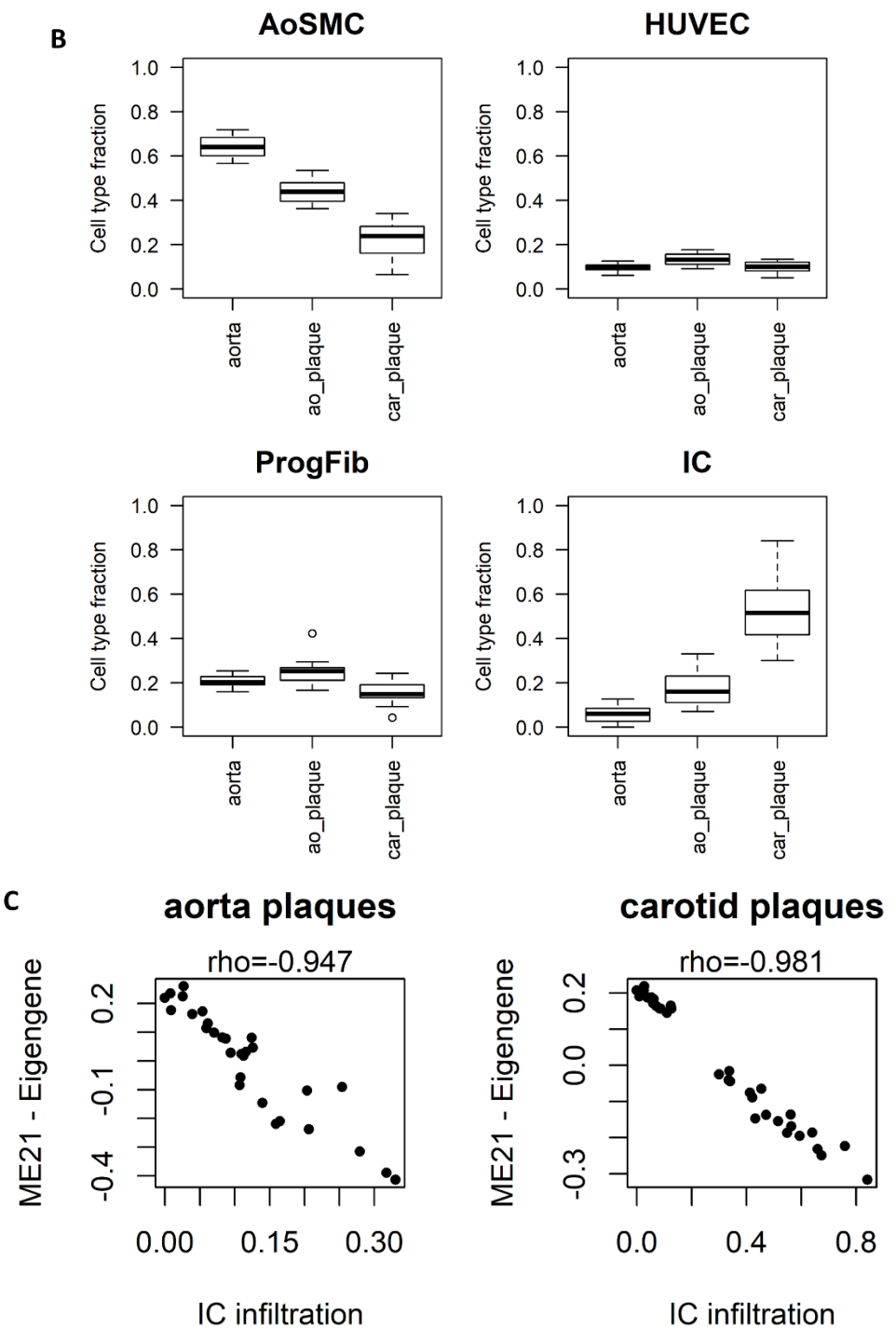
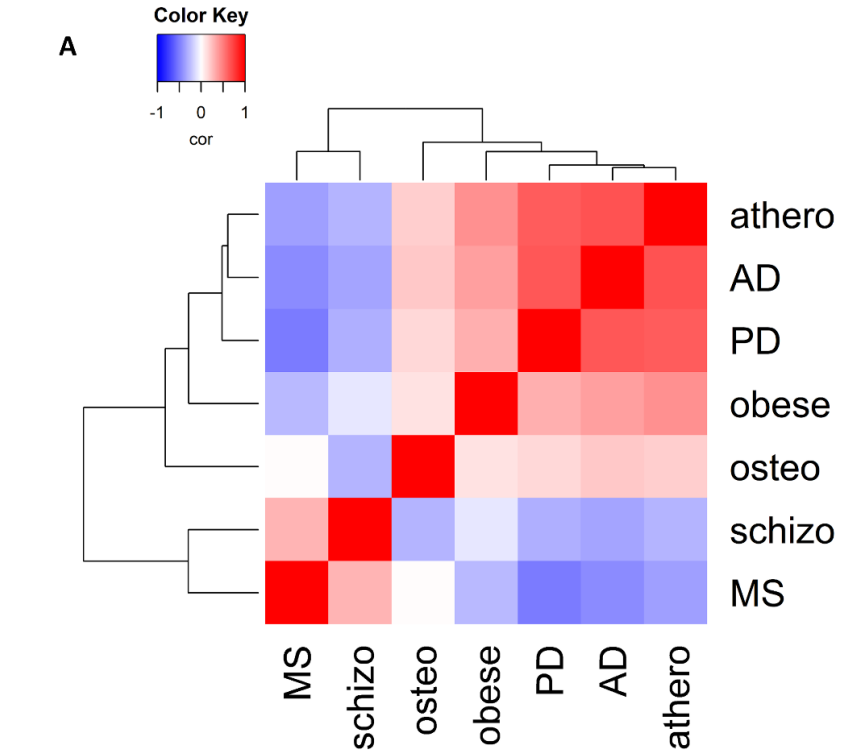


Figure 4: Immuno-methylation signature. A) Estimated blood immune cell type distribution shift in atherosclerosis and AD. **B)** Estimated cell type distribution in healthy aorta, aorta atherosclerotic plaque (ao_plaque) and carotid plaque (car_plaque). **C)** Correlation between immune cell infiltration and the eigengenes of the consensus module (ME21) in aorta and carotid plaques.

The consensus methylation module is also present in other chronic inflammaging diseases

Since inflammation is a common hallmark of many chronic aging diseases, we further checked whether the consensus immunomethylation module identified was also prevalent in other chronic inflammation and age-dependent diseases. We therefore reanalyzed Illumina 450K DNA methylation profiles of whole blood samples of Parkinson's disease (PD), schizophrenia, obesity, osteoporosis and multiple sclerosis (MS), and correlated the gene significance values of the CpG probes in module M21 across all the diseases (**Figure 5A**). A strong positive correlation could be observed with PD. Also obesity and osteoporosis showed a moderate positive correlation. On the other hand, schizophrenia and MS demonstrated a negative correlation.

The high correlation found in PD could again be attributed to a shift in CD4T and granulocyte blood levels. However, no such changes were detected in obesity and osteoporosis (**Figure 5B**). In osteoporosis, NK cell levels were slightly higher as compared to healthy samples. In obesity and schizophrenia, immune cell types didn't change dramatically. MS patients had opposite cell type distribution in comparison to atherosclerosis, AD and PD patients with higher CD4T- and B-cell levels and lower granulocyte levels.


A
Color Key
-1
0
1
cor
athero
AD
PD
obese
osteo
schizo
MS
MS
schizo
osteo
obese
PD
AD
athero

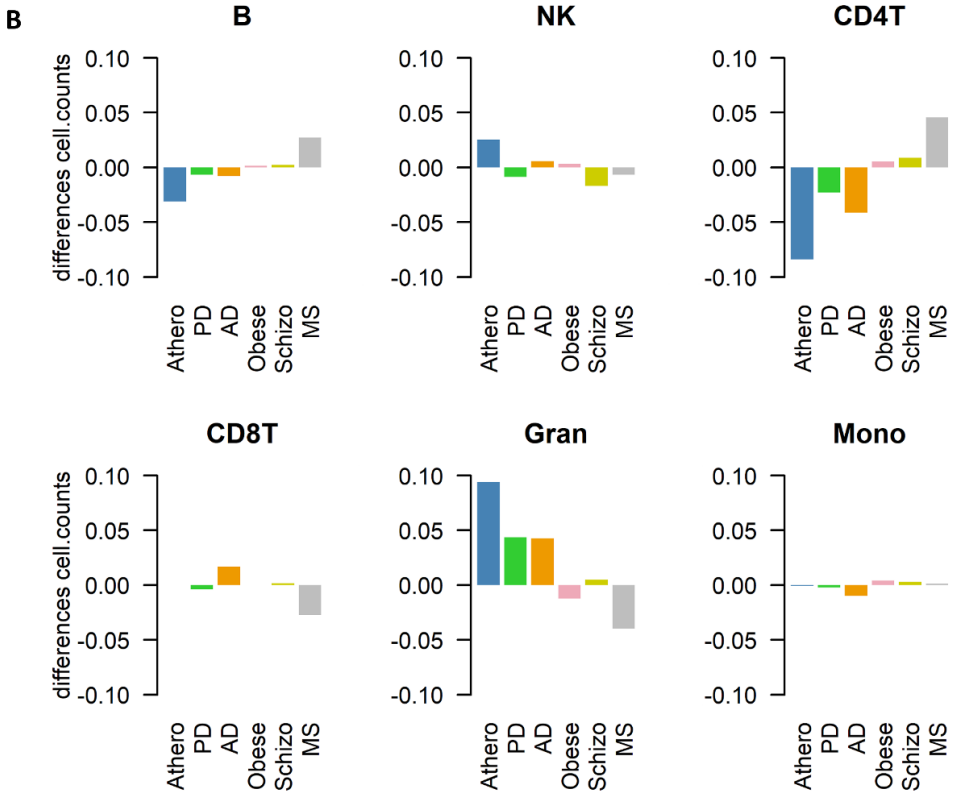

B
B
NK
CD4T
CD8T
Gran
Mono
differences cell.counts
0.10
0.05
0.00
-0.05
-0.10
Athero
PD
AD
Obese
Schizo
MS

Figure 5: The athero-AD consensus methylation module in whole blood of other inflamm-aging diseases. **A)** Correlation of module CpG probe gene significance values across six chronic diseases: atherosclerosis (athero), AD, PD, obesity (obese), osteoporosis (osteo), schizophrenia (schizo) and multiple sclerosis (MS). **B)** Estimated blood immune cell type distribution shifts in different chronic diseases.

Discussion

In this study, we identified a common DNA methylation signature in whole blood of atherosclerosis and AD patients. We showed that this consensus methylation module represents an immune component which correlates with shifts in blood immune cell distribution and immune cell infiltration in plaques and brains. Finally, we demonstrate the applicability of the immune-methylation signature, as an inflamm-aging disease biomarker.

Blood-based DNA methylation biomarkers can be valuable for diagnostic, predictive, prognostic and therapeutic purposes [41, 42]. Here, we showed that blood DNA methylation in atherosclerosis and AD are associated with similar shifts in immune cell type distribution and/or tissue infiltration. In both atherosclerosis and AD, granulocyte levels were increased while B and CD4T-cells were decreased. This is in accordance with other studies showing a higher neutrophil/lymphocyte ratio (NLR) in these diseases [43, 44]. NLR is a marker of systemic inflammation and has been found to be prognostic marker in CVDs associated with poor outcome and mortality [45, 46]. Interestingly, NLR can also be used to predict the presence of carotid atherosclerotic plaques [47]. Also in AD, NLR was higher as compared to healthy controls [43]. However, strong evidence for NLR as a prognostic or predictive biomarker in AD is lacking [48]. Whether our methylation profile is also a predictor of poor outcome or disease severity should be further investigated.

In cancer, systemic inflammation is associated with poor outcome [49, 50]. A recent study used DNA methylation to estimate NLR [51], and found that this methylation-derived NLR (mdNLR) was associated with poor survival in various cancer types [51, 52]. Furthermore, they also showed that mdNLR was increased with age [51]. Indeed, age is also accompanied by chronic low-level systemic inflammation, which is often called inflamm-aging [53]. In addition, many chronic diseases are more common with higher age, and it has therefore been suggested that aging and age-associated chronic diseases share the same underlying biological mechanisms [54, 55]. Many age-associated chronic diseases can therefore be seen as an acceleration of the aging process. Epigenetic clock age can be deduced from Illumina 450K DNA methylation profiles and accelerated epigenetic clock age has been associated with mortality and age-related diseases and phenotypes, suggesting that the epigenetic clock is a measure

for biological age, rather than chronological age [42, 56]. Interestingly, no single CpG site was in common between the immune-methylation signature identified in this study and the epigenetic clock signature, indicating a difference between the two DNA methylation-based biomarkers. Therefore, we also tested whether our methylation profile was present in other inflammation- and aging-associated diseases, besides AD and atherosclerosis. Remarkably, we observed a similar immunomethylation related change in PD. In contrast, the other diseases tested showed either low association with our methylation profile or no association, indicating that this profile is not a general marker for all inflamm-aging diseases. For example, MS showed a rather negative correlation with our methylation profile, which was also reflected in an opposite shift of cell type distribution, with higher lymphocytes and lower granulocytes levels. In contrast to MS, obesity showed a mild positive correlation, although this does not change the cell type contribution of the major blood cell types, which may indicate the involvement of other minority blood cell types or different activation cell activation states [57, 58].

In most blood-based EWAS, the Houseman algorithm is frequently applied to correct for variations in blood sample cell composition which may contribute to methylation variability [59]. However, we believe that this immune component may be an important determinant of aging disease etiologies and holds valuable information for prognostic or therapeutic biomarker applications. DNA methylation may be a very sensitive method to estimate small shifts in immune cell distribution or activation status. For example, a recent study found DNA methylation differences were associated with NK cell activation [60]. DNA methyltransferase DNMT3B seems to be important in regulating macrophage polarization [61]. In another study, FOXP3 methylation can be used to count regulatory T cells in blood and solid tissues [62]. A methylation CpG site in GPR15 gene which was associated with smoking, was found to be due to a higher proportion of CD3+GPR15+ expressing T cells in blood, and not by the direct effect of smoking on DNA methylation [63]. Correcting for cell type effects in EWAS is not always useful and may remove important information about the disease pathology [64]. In addition, even highly purified cell types were found to be rather a collection of epigenomes (which the authors called meta-epigenomes) [65], and may therefore not exclude all cellular effects. The usefulness of measuring cell type effects using DNA methylation was also exemplified by the extrinsic epigenetic clock which is influenced by blood cell counts. Faster extrinsic epigenetic age acceleration was associated with all-cause mortality [66], while different healthy lifestyle factors resulted in a decrease in extrinsic epigenetic age acceleration [67]. These results indicate that it may be useful to also include cellular effects which may be used to asses therapeutic, nutritional and lifestyle interventions.

Although this immune-associated DNA methylation profile is associated with atherosclerosis and AD, further longitudinal studies are required to estimate whether it is also related to disease outcome or progression. We established a correlation of DNA methylation changes with immune cell infiltration in atherosclerotic plaques and tumors. Immune cells play important roles in atherosclerosis and can either promoter or reduce atherosclerosis progression [68]. It would be interesting to study whether we can use blood-based methylation profiles to predict the inflammation status of atherosclerotic plaques. We showed that the hypermethylated profile in atherosclerotic plaques described previously [34], could be mainly attributed to increases in immune cells in the artery. This is of course not surprising as arteries and atherosclerotic plaques are a complex mixture of cell types and that atherosclerosis results in a dramatic remodeling of artery cell types, such as infiltration of immune cells and proliferation of smooth muscle cells. It is therefore questionable whether all the methylation changes detected in atherosclerotic plaques are due to intrinsic methylation changes in specific cell types and whether these aberrant DNA methylation marks could be targets for cell type specific therapeutic interventions. We also need to point out that our reference methylome-based estimation of the cell type counts in plaques could not be validated with histologically determined cell type counts and that the tissue consist of much more complex cell types which were not included in the reference methylome. Furthermore, we used ENCODE cell lines as reference methylomes which may not be completely representative for the cells *in vivo*. However, previous studies already used cell lines to estimate cell type fractions, and a recent study used the same ENCODE cell lines to estimate cell type counts in aortic samples in relation to ascending aortic dissection and bicuspid aortic valve [69, 70].

Due to the lack of brain cell type reference methylomes, we were unable to estimate immune cell infiltration in AD brain tissues or the contribution of microglia. However, neuro-inflammation plays an important role in AD and there is evidence that systemic immune cells may infiltrate into the brain [71]. Whether our methylation profile correlates with neuro-inflammation or number of infiltrated immune cells should be further investigated. Interestingly, we found no correlation of methylation in AD cerebellum samples with our immune-DNA methylation signature, which is in accordance with studies showing that the cerebellum is less susceptible to AD neuropathological features like amyloid plaques and neuronal loss than cortex and hippocampus [72].

In conclusion, inflammaging diseases, including atherosclerosis, AD, PD and obesity, share a common DNA methylation profiles in whole blood samples representing a disease-associated immune component reflected by changes in blood immune cell counts and predictive for immune cell infiltration in disease tissues. In addition to

epigenetic clock measurements, this immune-methylation signature may become a valuable blood-based biomarker to prevent chronic inflammatory disease development or monitor lifestyle intervention strategies which promote healthy aging.

References

1. Lathe, R., A. Sapronova, and Y. Kotelevtsev, *Atherosclerosis and Alzheimer--diseases with a common cause? Inflammation, oxysterols, vasculature.* BMC Geriatr, 2014. **14**: p. 36.
2. Ballard, C., et al., *Alzheimer's disease.* Lancet, 2011. **377**(9770): p. 1019-31.
3. Masters, C.L., et al., *Alzheimer's disease.* Nat Rev Dis Primers, 2015. **1**: p. 15056.
4. Kisler, K., et al., *Cerebral blood flow regulation and neurovascular dysfunction in Alzheimer disease.* Nat Rev Neurosci, 2017. **18**(7): p. 419-434.
5. Nelson, A.R., et al., *Neurovascular dysfunction and neurodegeneration in dementia and Alzheimer's disease.* Biochim Biophys Acta, 2016. **1862**(5): p. 887-900.
6. Zlokovic, B.V., *Neurovascular pathways to neurodegeneration in Alzheimer's disease and other disorders.* Nat Rev Neurosci, 2011. **12**(12): p. 723-38.
7. Casserly, I. and E. Topol, *Convergence of atherosclerosis and Alzheimer's disease: inflammation, cholesterol, and misfolded proteins.* Lancet, 2004. **363**(9415): p. 1139-46.
8. Lusis, A.J., *Atherosclerosis.* Nature, 2000. **407**(6801): p. 233-41.
9. Roher, A.E., et al., *Circle of willis atherosclerosis is a risk factor for sporadic Alzheimer's disease.* Arterioscler Thromb Vasc Biol, 2003. **23**(11): p. 2055-62.
10. Honig, L.S., W. Kukull, and R. Mayeux, *Atherosclerosis and AD: analysis of data from the US National Alzheimer's Coordinating Center.* Neurology, 2005. **64**(3): p. 494-500.
11. Roher, A.E., et al., *Atherosclerosis of cerebral arteries in Alzheimer disease.* Stroke, 2004. **35**(11 Suppl 1): p. 2623-7.
12. Dolan, H., et al., *Atherosclerosis, dementia, and Alzheimer disease in the Baltimore Longitudinal Study of Aging cohort.* Ann Neurol, 2010. **68**(2): p. 231-40.
13. Kim, T.W., et al., *Clinical effect of cerebrovascular atherosclerosis on cognition in Alzheimer's disease.* Arch Gerontol Geriatr, 2016. **63**: p. 55-8.
14. Arvanitakis, Z., et al., *Relation of cerebral vessel disease to Alzheimer's disease dementia and cognitive function in elderly people: a cross-sectional study.* Lancet Neurol, 2016. **15**(9): p. 934-943.
15. Roher, A.E., et al., *Intracranial atherosclerosis as a contributing factor to Alzheimer's disease dementia.* Alzheimers Dement, 2011. **7**(4): p. 436-44.
16. Yarchoan, M., et al., *Cerebrovascular atherosclerosis correlates with Alzheimer pathology in neurodegenerative dementias.* Brain, 2012. **135**(Pt 12): p. 3749-56.
17. Yuan, J., et al., *The occurrence of cerebrovascular atherosclerosis in Alzheimer's disease patients.* Clin Interv Aging, 2013. **8**: p. 581-4.
18. Silvestrini, M., et al., *Carotid atherosclerosis and cognitive decline in patients with Alzheimer's disease.* Neurobiol Aging, 2009. **30**(8): p. 1177-83.
19. Beeri, M.S., et al., *Coronary artery disease is associated with Alzheimer disease neuropathology in APOE4 carriers.* Neurology, 2006. **66**(9): p. 1399-404.
20. Hofman, A., et al., *Atherosclerosis, apolipoprotein E, and prevalence of dementia and Alzheimer's disease in the Rotterdam Study.* Lancet, 1997. **349**(9046): p. 151-4.
21. Silvestrini, M., et al., *The role of carotid atherosclerosis in Alzheimer's disease progression.* J Alzheimers Dis, 2011. **25**(4): p. 719-26.
22. Wendell, C.R., et al., *Carotid atherosclerosis and prospective risk of dementia.* Stroke, 2012. **43**(12): p. 3319-24.
23. Newman, A.B., et al., *Dementia and Alzheimer's disease incidence in relationship to cardiovascular disease in the Cardiovascular Health Study cohort.* J Am Geriatr Soc, 2005. **53**(7): p. 1101-7.
24. Kovacic, J.C. and V. Fuster, *Atherosclerotic risk factors, vascular cognitive impairment, and Alzheimer disease.* Mt Sinai J Med, 2012. **79**(6): p. 664-73.
25. Fiolaki, A., et al., *Atherosclerosis, biomarkers of atherosclerosis and Alzheimer's disease.* Int J Neurosci, 2014. **124**(1): p. 1-11.
26. Zhu, H., et al., *The association of apolipoprotein E (APOE) gene polymorphisms with atherosclerosis susceptibility: a meta-analysis.* Minerva Cardioangiol, 2016. **64**(1): p. 47-54.

27. Yu, J.T., L. Tan, and J. Hardy, *Apolipoprotein E in Alzheimer's disease: an update.* Annu Rev Neurosci, 2014. **37**: p. 79-100.
28. Istas, G., et al., *Identification of differentially methylated BRCA1 and CRISP2 DNA regions as blood surrogate markers for cardiovascular disease.* Sci Rep, 2017. **7**(1): p. 5120.
29. Zaina, S., et al., *DNA methylation dynamics in human carotid plaques after cerebrovascular events.* Arterioscler Thromb Vasc Biol, 2015. **35**(8): p. 1835-42.
30. Nazarenko, M.S., et al., *A comparison of genome-wide DNA methylation patterns between different vascular tissues from patients with coronary heart disease*. PLoS One, 2015. **10**(4): p. e0122601.
31. Valencia-Morales Mdel, P., et al., *The DNA methylation drift of the atherosclerotic aorta increases with lesion progression.* BMC Med Genomics, 2015. **8**: p. 7.
32. Zaina, S., *Unraveling the DNA methylome of atherosclerosis.* Curr Opin Lipidol, 2014. **25**(2): p. 148-53.
33. Yamada, Y., et al., *Identification of hypo- and hypermethylated genes related to atherosclerosis by a genome-wide analysis of DNA methylation.* Int J Mol Med, 2014. **33**(5): p. 1355-63.
34. Zaina, S., et al., *DNA methylation map of human atherosclerosis.* Circ Cardiovasc Genet, 2014. **7**(5): p. 692-700.
35. De Jager, P.L., et al., *Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci.* Nat Neurosci, 2014. **17**(9): p. 1156-63.
36. Lunnon, K., et al., *Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease.* Nat Neurosci, 2014. **17**(9): p. 1164-70.
37. Mano, T., et al., *Neuron-specific methylome analysis reveals epigenetic regulation and tau-related dysfunction of BRCA1 in Alzheimer's disease.* Proc Natl Acad Sci U S A, 2017. **114**(45): p. E9645-E9654.
38. Yu, L., et al., *Methylation profiles in peripheral blood CD4+ lymphocytes versus brain: The relation to Alzheimer's disease pathology.* Alzheimers Dement, 2016. **12**(9): p. 942-951.
39. Li, H., et al., *Common DNA methylation alterations of Alzheimer's disease and aging in peripheral whole blood.* Oncotarget, 2016. **7**(15): p. 19089-98.
40. Thorsson, V., et al., *The Immune Landscape of Cancer.* Immunity, 2018. **48**(4): p. 812-830 e14.
41. Berdasco, M. and M. Esteller, *Clinical epigenetics: seizing opportunities for translation*. Nat Rev Genet, 2019. **20**(2): p. 109-127.
42. Horvath, S. and K. Raj, *DNA methylation-based biomarkers and the epigenetic clock theory of ageing.* Nat Rev Genet, 2018. **19**(6): p. 371-384.
43. Kuyumcu, M.E., et al., *The evaluation of neutrophil-lymphocyte ratio in Alzheimer's disease.* Dement Geriatr Cogn Disord, 2012. **34**(2): p. 69-74.
44. Balta, S., et al., *The Relation Between Atherosclerosis and the Neutrophil-Lymphocyte Ratio.* Clin Appl Thromb Hemost, 2016. **22**(5): p. 405-11.
45. Teperman, J., et al., *Relationship between neutrophil-lymphocyte ratio and severity of lower extremity peripheral artery disease.* Int J Cardiol, 2017. **228**: p. 201-204.
46. Xue, J., et al., *Neutrophil-to-Lymphocyte Ratio Is a Prognostic Marker in Acute Ischemic Stroke.* J Stroke Cerebrovasc Dis, 2017. **26**(3): p. 650-657.
47. Corriere, T., et al., *Neutrophil-to-Lymphocyte Ratio is a strong predictor of atherosclerotic carotid plaques in older adults.* Nutr Metab Cardiovasc Dis, 2018. **28**(1): p. 23-27.
48. Rembach, A., et al., *An increased neutrophil-lymphocyte ratio in Alzheimer's disease is a function of age and is weakly correlated with neocortical amyloid accumulation.* J Neuroimmunol, 2014. **273**(1-2): p. 65-71.
49. Diakos, C.I., et al., *Cancer-related inflammation and treatment effectiveness.* Lancet Oncol, 2014. **15**(11): p. e493-503.
50. Rossi, S., et al., *Are Markers of Systemic Inflammation Good Prognostic Indicators in Colorectal Cancer?* Clin Colorectal Cancer, 2017. **16**(4): p. 264-274.
51. Koestler, D.C., et al., *DNA Methylation-Derived Neutrophil-to-Lymphocyte Ratio: An Epigenetic Tool to Explore Cancer Inflammation and Outcomes.* Cancer Epidemiol Biomarkers Prev, 2017. **26**(3): p. 328-338.

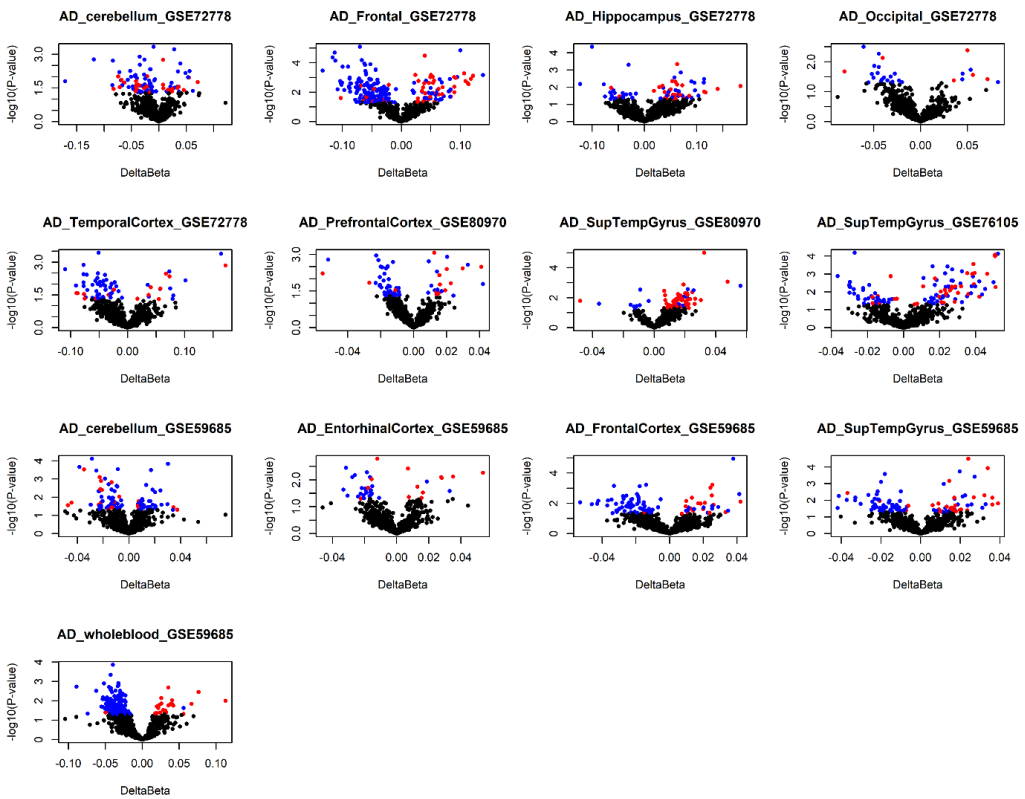
52. Wiencke, J.K., et al., *Immunomethylomic approach to explore the blood neutrophil lymphocyte ratio (NLR) in glioma survival.* Clin Epigenetics, 2017. **9**: p. 10.
53. Franceschi, C., et al., *Inflammaging: a new immune-metabolic viewpoint for age-related diseases.* Nat Rev Endocrinol, 2018. **14**(10): p. 576-590.
54. Franceschi, C., et al., *The Continuum of Aging and Age-Related Diseases: Common Mechanisms but Different Rates.* Front Med (Lausanne), 2018. **5**: p. 61.
55. Kennedy, B.K., et al., *Geroscience: linking aging to chronic disease.* Cell, 2014. **159**(4): p. 709-13.
56. Declerck, K. and W. Vanden Berghe, *Back to the future: Epigenetic clock plasticity towards healthy aging.* Mech Ageing Dev, 2018. **174**: p. 18-29.
57. DeFuria, J., et al., *B cells promote inflammation in obesity and type 2 diabetes through regulation of T-cell function and an inflammatory cytokine profile.* Proc Natl Acad Sci U S A, 2013. **110**(13): p. 5133-8.
58. Touch, S., K. Clement, and S. Andre, *T Cell Populations and Functions Are Altered in Human Obesity and Type 2 Diabetes.* Curr Diab Rep, 2017. **17**(9): p. 81.
59. Jaffe, A.E. and R.A. Irizarry, *Accounting for cellular heterogeneity is critical in epigenome-wide association studies.* Genome Biol, 2014. **15**(2): p. R31.
60. Wiencke, J.K., et al., *The DNA methylation profile of activated human natural killer cells.* Epigenetics, 2016. **11**(5): p. 363-80.
61. Yang, X., et al., *Epigenetic regulation of macrophage polarization by DNA methyltransferase 3b.* Mol Endocrinol, 2014. **28**(4): p. 565-74.
62. Wieczorek, G., et al., *Quantitative DNA methylation analysis of FOXP3 as a new method for counting regulatory T cells in peripheral blood and solid tissue.* Cancer Res, 2009. **69**(2): p. 599-608.
63. Bauer, M., et al., *A varying T cell subtype explains apparent tobacco smoking induced single CpG hypomethylation in whole blood.* Clin Epigenetics, 2015. **7**: p. 81.
64. Holbrook, J.D., et al., *Is cellular heterogeneity merely a confounder to be removed from epigenome-wide association studies?* Epigenomics, 2017. **9**(8): p. 1143-1150.
65. Wijetunga, N.A., et al., *The meta-epigenomic structure of purified human stem cell populations is defined at cis-regulatory sequences.* Nat Commun, 2014. **5**: p. 5195.
66. Chen, B.H., et al., *DNA methylation-based measures of biological age: meta-analysis predicting time to death.* Aging (Albany NY), 2016. **8**(9): p. 1844-1865.
67. Quach, A., et al., *Epigenetic clock analysis of diet, exercise, education, and lifestyle factors.* Aging (Albany NY), 2017. **9**(2): p. 419-446.
68. Hansson, G.K. and P. Libby, *The immune response in atherosclerosis: a double-edged sword.* Nat Rev Immunol, 2006. **6**(7): p. 508-19.
69. Pan, S., et al., *DNA methylome analysis reveals distinct epigenetic patterns of ascending aortic dissection and bicuspid aortic valve.* Cardiovasc Res, 2017. **113**(6): p. 692-704.
70. Zheng, S.C., et al., *A novel cell-type deconvolution algorithm reveals substantial contamination by immune cells in saliva, buccal and cervix.* Epigenomics, 2018. **10**(7): p. 925-940.
71. Prinz, M. and J. Priller, *The role of peripheral immune cells in the CNS in steady state and disease.* Nat Neurosci, 2017. **20**(2): p. 136-144.
72. Heneka, M.T., et al., *Neuroinflammation in Alzheimer's disease.* Lancet Neurol, 2015. **14**(4): p. 388-405.

Supplementary information

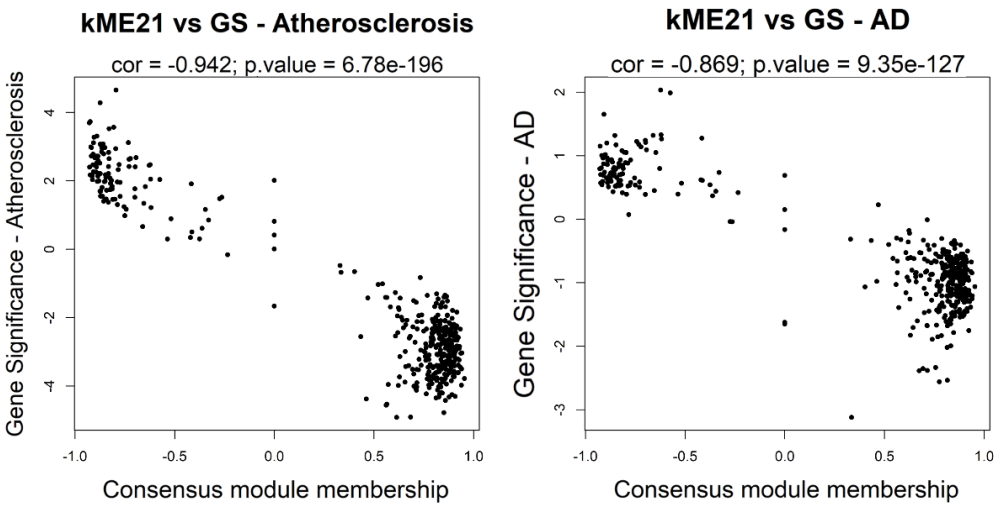
Supplementary tables 1 and 2 can be found in following dropbox folder: https://www.dropbox.com/sh/d9lq9mipqkoezx3/AAC9RCXgWS7P6DS71ybJqWNJa?dl=0

Supplementary Table 1: Illumina CpG probes used as reference methylome to estimate cell type counts in aorta healthy tissue and atherosclerotic plaques.

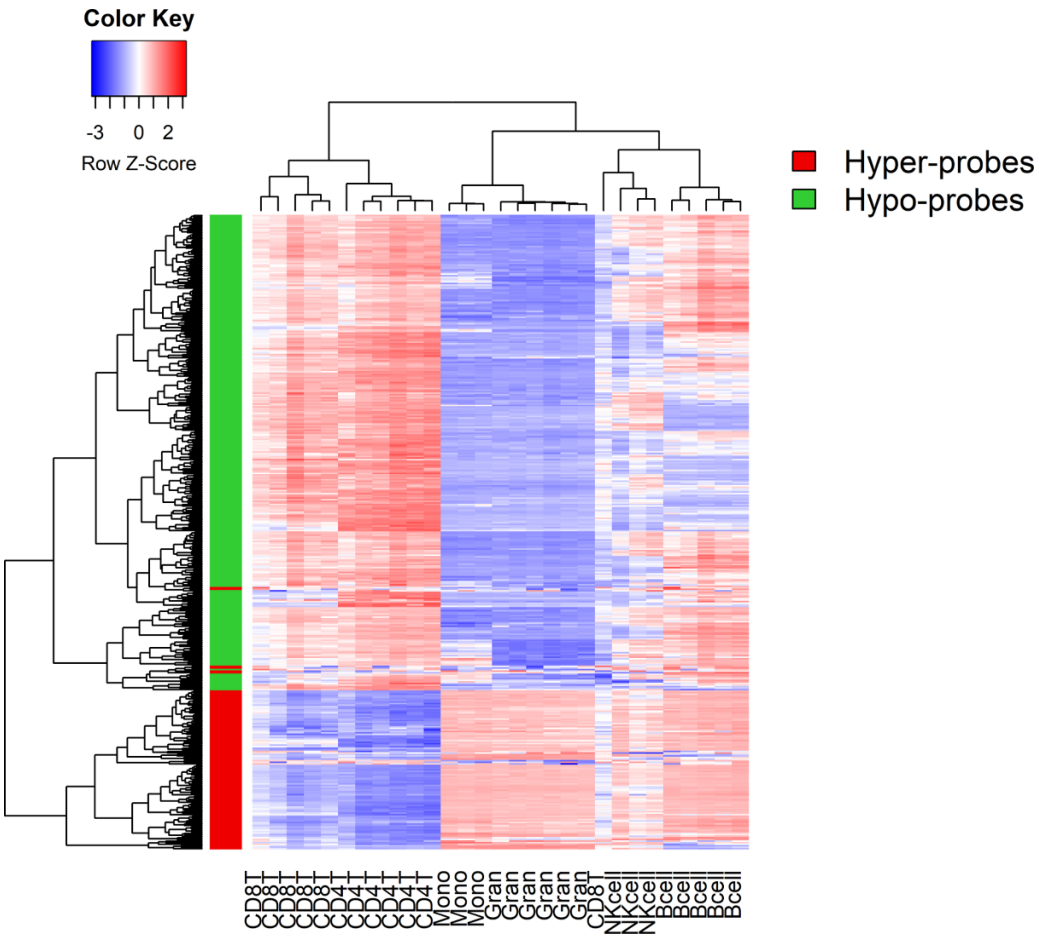
Supplementary Table 2: Illumina CpG probes of consensus module 21. For each probe the Gene Significance value (GS) and Module Membership (kME21) is given.



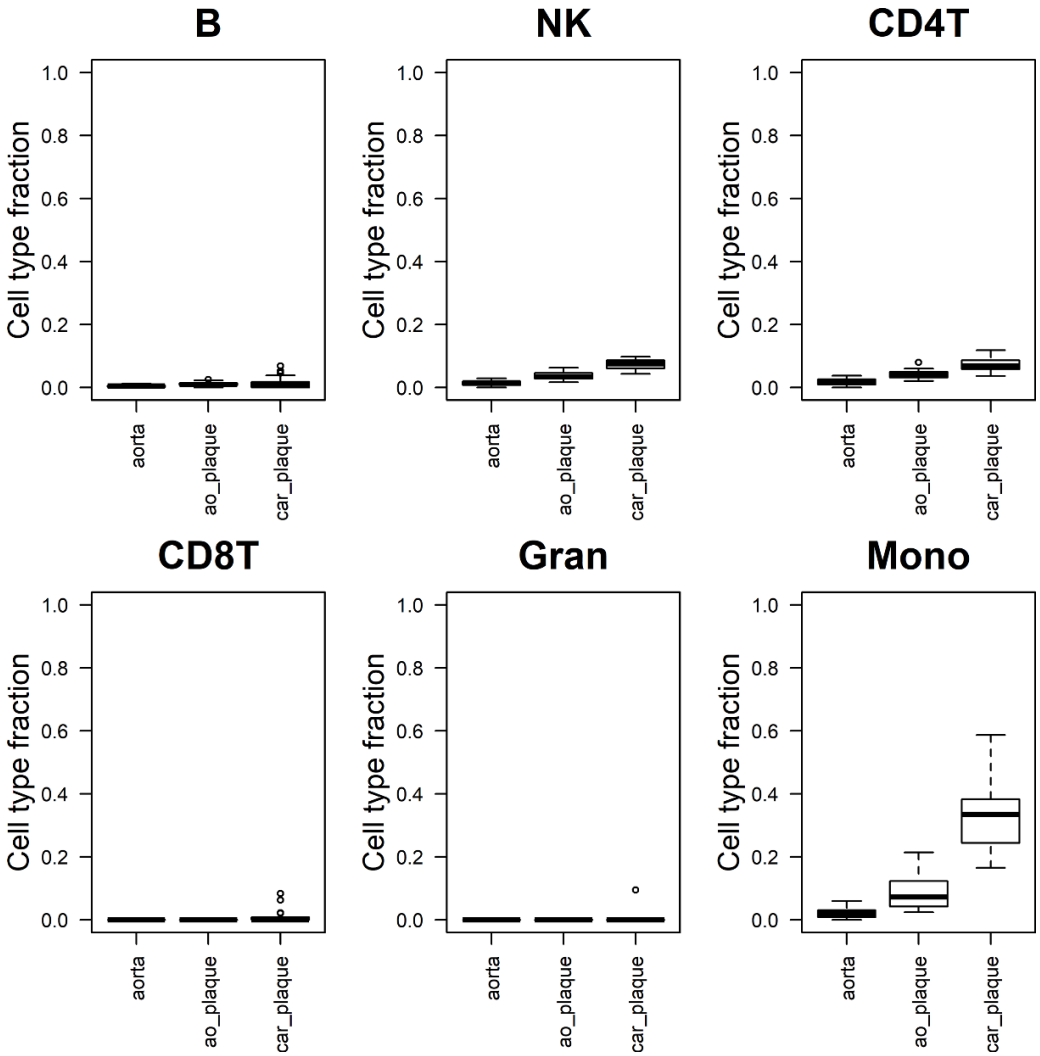
Supplementary Figure 1: DNA methylation similarity between atherosclerosis whole blood samples and AD brain tissues. Volcano plots showing the methylation differences of the 712 top significant atherosclerosis DMPs (athero-DMPs) in different publicly available AD brain datasets (see Table 1). Significant probes (p-value < 0.05) were colored blue when hypomethylated and red when hypermethylated in atherosclerosis.



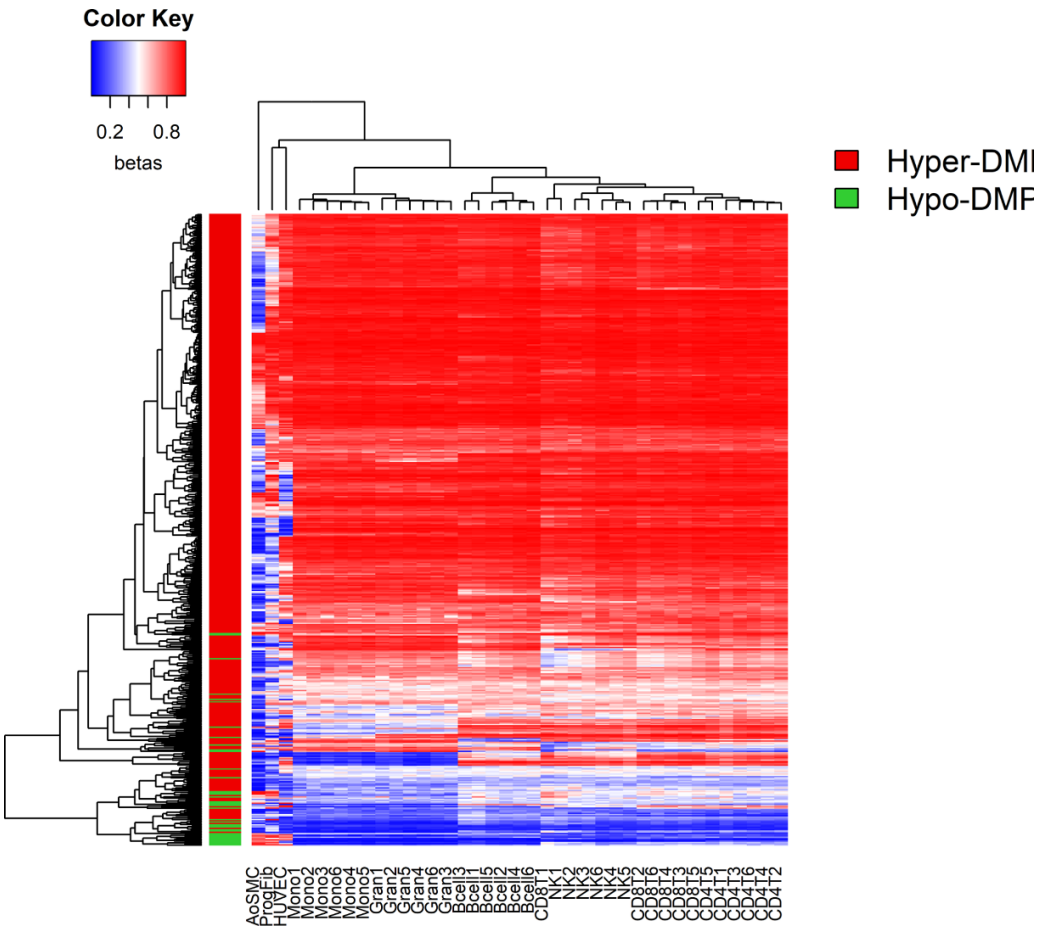
Supplementary Figure 2: Correlation between gene significance and module membership of probes in module 21 in atherosclerosis (left) and AD (right).



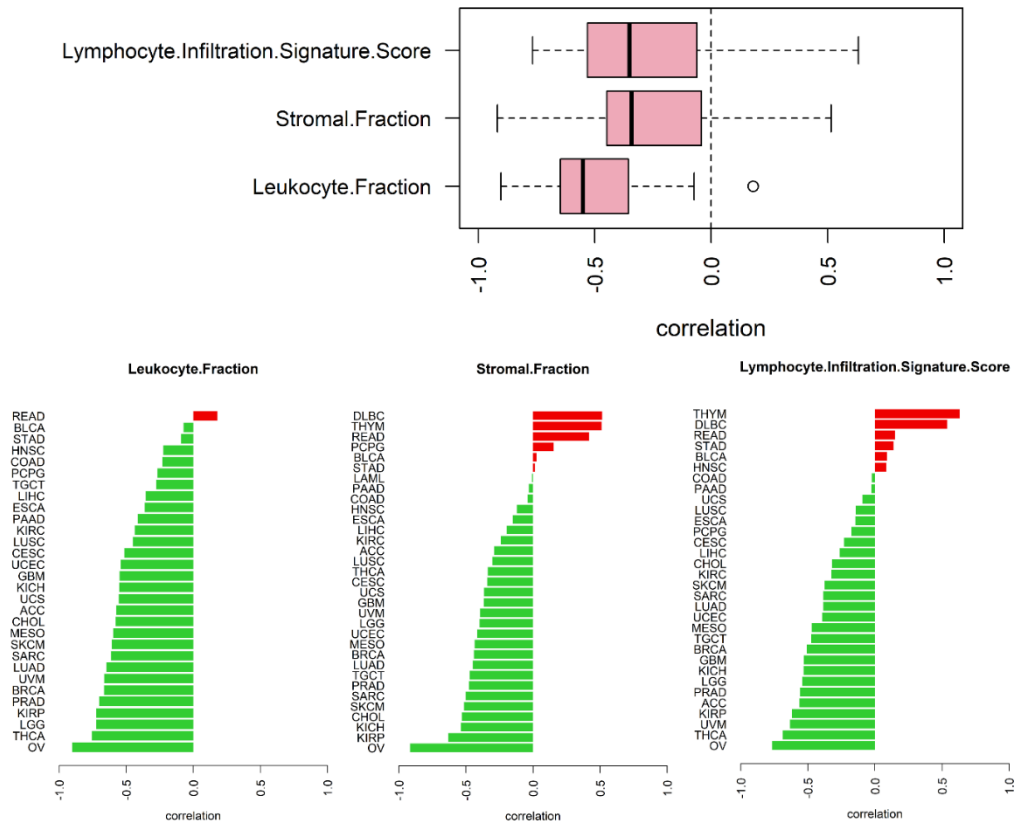
Supplementary Figure 3: DNA methylation values of the consensus module CpG probes across the major blood immune cell types.



Supplementary Figure 4: Estimated immune cell fraction in healthy aorta, aorta atherosclerotic plaques (ao_plaque) and carotid plaques (car_plaque).



Supplementary Figure 5: DNA methylation levels of the top 500 most significant differentially methylated probes in aorta atherosclerotic plaques compared to healthy aorta tissues across different vascular cell types (AoSMC: aorta smooth muscle cells, ProgFib: fibroblasts, HUVEC: endothelial cells).



Supplementary Figure 6: TCGA immune cell infiltration. Correlation between the consensus module eigengenes of TCGA cancers and measures of immune cell infiltration (Leukocyte fraction, stromal fraction and lymphocyte infiltration signature score).

7

Strong interindividual epigenetic variation in response to flavanol diet intervention reveals redundant gene wiring of cell adhesion networks

Ken Declerck[1], Dragan Milenkovic[2], Katarzyna Szarc vel Szic[1], Clarissa Gerhauser[3], Guido R. M. M. Haenen[4], Aalt Bast[4], Antje R. Weseler[4], Christine Morand[2], Wim Vanden Berghe[1]

[1] Laboratory of Protein Chemistry, Proteomics and Epigenetic Signaling (PPES), Department of Biomedical Sciences, University of Antwerp, Belgium
[2] INRA, UMR 1019, UNH, CRNH Auvergne, F-63000 Clermont-Ferrand; Clermont Université, Université d'Auvergne, Unité de Nutrition Humaine, BP 10448, F-63000, Clermont-Ferrand, France
[3] Division of Epigenomics and Cancer Risk Factors, DKFZ, Heidelberg, Germany
[4] Department of Toxicology, Maastricht University, Maastricht, The Netherlands

Adapted from:

- *Dietary Flavanols Modulate the Transcription of Genes Associated with Cardiovascular Pathology without Changes in Their DNA Methylation State. Dragan Milenkovic, Wim Vanden Berghe, Céline Boby, Christine Leroux, **Ken Declerck**, Katarzyna Szarc vel Szic, Karen Heyninck, Kris Laukens, Martin Bizet, Matthieu Defrance, Sarah Dedeurwaerder, Emilie Calonne, Francois Fuks, Guy Haegeman, Guido R. M. M. Haenen, Aalt Bast, Antje R. Weseler, PLoS One. 2014; 9(4): e95527.*
- *A systems biology network analysis of nutri(epi)genomic changes in endothelial cells exposed to epicatechin metabolites. Milenkovic D, Vanden Berghe WV, Morand C, Claude S, van de Sandt A, Gorressen S, Monfoulet LE, Chirumamilla CS, **Declerck K**, Szic*

KSV, Lahtela-Kakkonen M, Gerhauser C, Merx MW, Kelm M. Sci Rep. 2018 Oct 19;8(1):15487.

- *Epicatechin metabolites exert epigenetic regulation in endothelial cells by modulating DNA methylation profile.* ***Ken Declerck****(*), Dragan Milenkovic(*), Yelena Guttman, Zohar Kerem, Sylvain Claude, Hagen Schroeter, Christine Morand, Wim Vanden Berghe (*equally contributed), submitted in Biochemical Pharmacology.*

Contributions: Ken Declerck analyzed the DNA methylation data, assisted in interpreting the results and writing of the papers.

Abstract | Diets rich in epicatechin flavanols are known to exert cardioprotective effects through reduction of monocyte-endothelial cell adhesion and transendothelial monocyte migration. In the present study we apply genome-wide DNA methylation profiling to examine potential involvement of epigenetic remodeling in response to epicatechin metabolites in *in vitro* and *in vivo* studies. In line with their known biological properties, flavanol metabolites revealed endothelial DNA methylation changes in genes involved in cell adhesion, cytoskeletal organization, renin angiotensin, nitric oxide and axonal guidance signaling pathways. Upon flavanol enriched diet intervention, we observed a strong interindividual variation in blood DNA gene methylation response which affects common cell adhesion pathways. The high interindividual gene diversity suggests highly redundant epigenetic wiring of cell adhesion networks, beyond direct transcriptional regulation. Furthermore, we identified reciprocal atherosclerosis and flavanol diet specific epigenetic changes, which involve systemic changes in blood cell types.

Introduction

A diet rich in fruits and vegetables is known to have beneficial health effects partly due the high levels of polyphenolic flavonoids [1]. Consumption of flavanols, a subclass of flavonoids, has been associated with improved cardiovascular health [2-5]. The main flavanol compounds are the monomeric catechin and epicatechin and the oligomeric procyanidins with cocoa, apples and tea as the major food sources. Only monomers and dimers are absorbed in the small intestine, and undergo metabolization in the liver leading to glucorinidated, sulphated and methylated metabolites [6].

Experimental studies and randomized controlled trials (RCTs) further confirm the cardioprotective effects of flavanols [7] including effects on blood pressure [8, 9], vasodilation [10, 11], and LDL cholesterol [12]. The most consistent effect is on endothelial function measured by flow-mediated vasodilation (FMD) [13], which is an independent risk factor for CVDs. Multiple RCTs showed an increase in FMD and reduction in blood pressure, both in CVD high risk and low risk individuals [14, 15]. This effect is mainly due to the enhancement of NO synthesis and levels in the vascularity [10, 16]. Furthermore, (-)-epicatechin was found to partially cause this increase in FMD after the consumption of cocoa flavanol-rich diet [7].

In a previous study, a four and eight week diet intervention with monomeric and oligomeric flavanols (MOF) from grape seeds in 28 male smokers revealed pleiotropic vascular health benefits [17]. Although only a few macro- and microvascular biomarkers were significantly improved after the diet intervention, integrating all

measures into a vascular health index resulted in an overall improvement in vascular health by MOF diet supplementation, which was not seen in the placebo group [17]. In an attempt to better understand molecular mechanisms of flavanols, genome-wide gene expression levels were measured before and after the diet intervention in blood samples which reveals expression changes in genes involved in chemotaxis, cell adhesion, cell infiltration and cytoskeleton organization [18]. In addition, *in vitro* assays demonstrated that MOF exposure of monocytes reduces cell-cell adhesion to endothelial cells and decreases LPS-stimulated TNF-α expression in macrophages [18]. Also in endothelial cells, a mixture of flavanol metabolites induced transcriptomic changes in cell adhesion genes [19]. Therefore, a possible alternative way how flavanols may have beneficial effects on endothelial function and prevent CVD is by reducing the recruitment and migration of monocytes into the arterial wall.

In the previous part of this thesis, we identified an atherosclerosis-associated DNA methylation profile in blood. It would be interesting to know whether therapeutic or nutritional interventions, including flavanols could reverse this aberrant atherosclerosis methylation profile back to normal levels, and whether blood DNA methylation profiles could be used as a dynamic health biomarker. Polyphenolic compounds were found to induce DNA methylation changes, mainly in *in vitro* settings [20]. How these compounds and other environmental factors may affect epigenetic and DNA methylation profiles is not completely understood, but may involve changes in cellular metabolism and/or direct interfering with epigenetic enzymes including DNMTs. For example, EGCG, the major polyphenol from green tea, can inhibit DNMT activity through the formation of hydrogen bounds in the catalytic pocket of DNMT resulting in DNA methylation changes in cancer cell lines [21] and the re-activation of silenced tumor suppressor genes [22, 23]. In addition, DNMT activity is dependent on intracellular SAM levels, which is the main substrate for DNMTs. Accordingly, changes in cellular metabolism which affect methyl-donor availability will also be reflected at the level of DNA methylation [24].

The prognostic value of blood DNA methylation markers was recently being demonstrated by the development of a multi-tissue epigenetic clock which estimates chronological age [25, 26]. This clock was also found to be a marker for biological age as it could predict all-cause and cardiovascular mortality [27]. In addition, accelerated epigenetic ageing was found to correlate with multiple age-related disorders and phenotypes [28]. Interesting, healthy lifestyle factors, including fish intake, moderate alcohol consumption, education and fruit/vegetables intake, were found to be associated with a younger epigenetic age than chronological age [29].

In this study, we have investigated the impact of flavanols on DNA methylation. First, provide proof of concept for DNA methylation effects *in vitro*, upon treatment of

endothelial HUVECs with different flavanol metabolites. Next, we characterized DNA methylation changes in an eight-week grape seed flavanol-rich diet intervention in blood samples of male smokers. Finally, we explored whether the DNA methylation effects by the grape seed diet intervention are correlated with vascular health markers and whether it can reverse or modulate atherosclerosis- and smoking-associated DNA methylation profiles.

Materials and methods

HUVEC *in vitro* treatments

Purified epicatechin metabolites: 3'-O-methyl(-)-epicatechin (3'MEC), 4'-O-methyl(-)-epicatechin-7-β-D-glucuronide (4'MEC7G) and (-)-epicatechin-4'-sulfate (EC4'S) were gifted by Mars Inc (McLean, VA, USA). Conjugated metabolites were dissolved in 50% ethanol in double distilled water. TNF-α was obtained from R&D Systems (Lille, France) and was dissolved in 0.1% BSA/P.

HUVECs (Lonza, Walkersville, MD, USA), a pool from four donors, at passage 3 were cultured in a phenol-free endothelial growth medium (EGM) supplemented with 2% fetal bovine serum (FBS), 0.4% fibroblast growth factor, 0.1% vascular endothelial growth factor, 0.1% heparin, 0.1% insulin-like growth factor, 0.1% ascorbic acid, 0.1% epidermal growth factor and 0.04% hydrocortisone (all from Lonza).

HUVECs, at passage 4, were seeded into the 24-well plates (BD-Falcon, Le Pont-De-Claix, France) and grown to reach 60-70% confluence. The medium was then replaced to expose the cells for 3 h to the culture medium containing either solvent alone (ethanol 0.5‰, control wells, vehicle), different metabolites: 3'MEC, 4'MEC7G or EC4'S or mixture of the metabolites at 1 μM. At the end of the incubation period, the medium was replaced with flavanol free medium and left for further 18 hours until confluence. The confluent monolayer was stimulated for 4 hours with TNF-α (R&D Systems Lille, France) at 0.1 ng/mL or only incubated with PBS/BSA (0,01‰, negative control).

Grape seed diet intervention

Details on the design, protocol and conduction of the clinical study are provided elsewhere [17]. In brief, non-obese, healthy male smokers with age between 30 and 60 years and smoking 10 and more cigarettes per day for at least five years were included in the trial. The study was approved by the Medical Ethical Committee of the Maastricht University and Academic Hospital Maastricht, The Netherlands (ClinicalTrials.gov, NCT00742287) and conducted according to the World Medical Association Declaration

of Helsinki of 1975, revision 2008. All subjects provided their written informed consent to participate in this study. For the present investigations blood samples of 13 men were available before and after an intervention of eight weeks with a daily intake of two capsules containing 100 mg MOF from Vitis vinifera L. seeds (MASQUELIER's Original OPCs, INC BV, Loosdrecht, The Netherlands). The composition of the MOF capsules is specified in [17]. Basically, the intake of 2 capsules per day provided each subject with 51.3 mg total catechins ((+)−catechin, (−)-epicatechin and (−)−epicatechin-3-O-gallate), 55 mg total flavanol dimers (proanthocyanidin B1, B2, B3 and B4) and 93.8 mg total tri-, tetra- and pentameric proanthocyanidins. During the intervention period subjects maintained their usual smoking, dietary and lifestyle habits, complied with the regular intake of the capsules and did not experience any severe side effects.

Genomic DNA extraction and bisulfite conversion

gDNA from HUVEC and EDTA-treated blood was extracted using DNeasy Blood & Tissue kit (Qiagen, Courtaboeuf, France). DNA purity and concentrations were determined by UV-VIS spectrophotometry (NanoDrop, Thermo Fisher Scientific Inc, Wilmington, DE, USA) and stored at -80°C until further use.

For each sample 1 µg of gDNA was bisulfite converted using the EZ DNA Methylation kit (Zymo research, Irvine, CA, USA) according to manufacturer's instruction. Successful bisulfite conversion was confirmed by performing a methylation-sensitive PCR in a region of the SALL3 gene. (see [30] for primer sequences).

Illumina DNA methylation arrays

DNA methylation levels were measured using the Illumina Infinium HD Methylation arrays (Illumina, San Diego, CA, USA) according to the manufacturer's instructions. Briefly, 4 µL (150 ng) of bisulfite-converted DNA was isothermally amplified overnight (20–24h) and fragmented enzymatically. Precipitated DNA was resuspended in hybridization buffer and dispensed onto the Infinium HumanMethylation450 BeadChips (12 samples/chip) using a Freedom EVO robot (Tecan, Männedorf, Switzerland). The hybridization procedure was performed at 48°C overnight (16–20 h) using an Illumina Hybridization oven. After hybridization, free DNA was washed away and the BeadChips were processed through a single nucleotide extension followed by fluorescent readout of the incorporated base using a Freedom EVO robot. Finally, the BeadChips were imaged using an Illumina iScan (Illumina, San Diego, CA, USA). Illumina data have been deposited in accordance to the MIAME guidelines in the public GEO database (http://www.ncbi.nlm.nih.gov/geo/), accession number GSE54690 (diet intervention study) and *GSEXXX* (*in vitro* data).

DNA methylation analysis

Filtering of bad quality probes and normalization of raw methylation beta values was conducted using the R package RnBeads [31]. Probes with detection p-values higher than 0.01, overlapping with SNPs at the last 3 bases in its sequence or containing missing values were excluded. BMIQ was used to normalize between the two different probe designs (Infinium I and Infinium II) [32]. Normalized methylation beta-values were logit-transformed into M-values. DMPs were identified using moderated t-test in the limma R package [33]. The Benjamini-Hochberg FDR was used to control for multiple testing. Variable methylated positions (VMPs) were identified using the Barlett's test in the iEVORA R script [34]. Probes were assigned to genes, CGI annotations (CGI, CGI shores, CGI shelves, and open sea) and gene regions (TSS1500, TSS200, 5'UTR, 1st exon, gene body and 3'UTR) based on the HumanMethylation450 v1.2 Manifest file from Illumina. Probes were also mapped to chromatin segmentation states obtained from the UCSC genome browser. Blood cell type composition was estimated using the reference-based method of Houseman with the EpiDISH R package [35, 36]. All analyses were performed in R and R Studio. Epigenetic age was estimated using the epigenetic clock of Horvath [25].

Pathway analysis

DMPs which mapped to genes according the HumanMethylation450 v1.2 Manifest file from Illumina were used to perform pathway enrichment analysis using the IPA software. Canonical pathways were said to be significant overrepresented when they had a Fisher's Exact p-value < 0.05.

Results

Flavanol metabolites modify DNA methylation profiles in endothelial cells

Genome-wide DNA methylation profiling of TNF-stimulated HUVECs treated with flavanol metabolites were performed. Using the DNA methylation values of all CpG probes on the array, we performed PCA (**Figure 1A**). DNA methylation profiles of the HUVECs treated with a mixture of flavanol metabolites (MIX) clearly formed a separate cluster, indicating that the methylation profile of these cells are different from the other samples. In addition, HUVECs treated with single flavanol metabolites also demonstrated a different profile compared to TNF-α and untreated conditions,

however, less pronounced as in the MIX condition. We also noticed that one control sample clustered together with the flavanol metabolites rather than with the other control samples.

We next selected CpG probes which were significantly differentially methylated between flavanol treatment and TNF-α treatment based on a p-value < 0.01 and a |delta beta| > 0.05 (**Figure 1B**). As expected, MIX treatment induced the strongest methylation changes leading to 22,491 hyper-DMPs and 5,666 hypo-DMPs. The number of DMPs induced by single flavanol metabolites were much lower: 313 hyper-DMPs and 393 hypo-DMPs in 3'MEC condition, 902 hyper-DMPs and 740 hypo-DMPs in 4'MEC7G condition and, 878 hyper-DMPs and 548 hypo-DMPs in EC4'S condition. TNF-α also demonstrated a change in methylation compared to untreated control samples resulting in 559 hyper-DMPs and 198 hypo-DMPs (**Supplementary Table 1-5**).

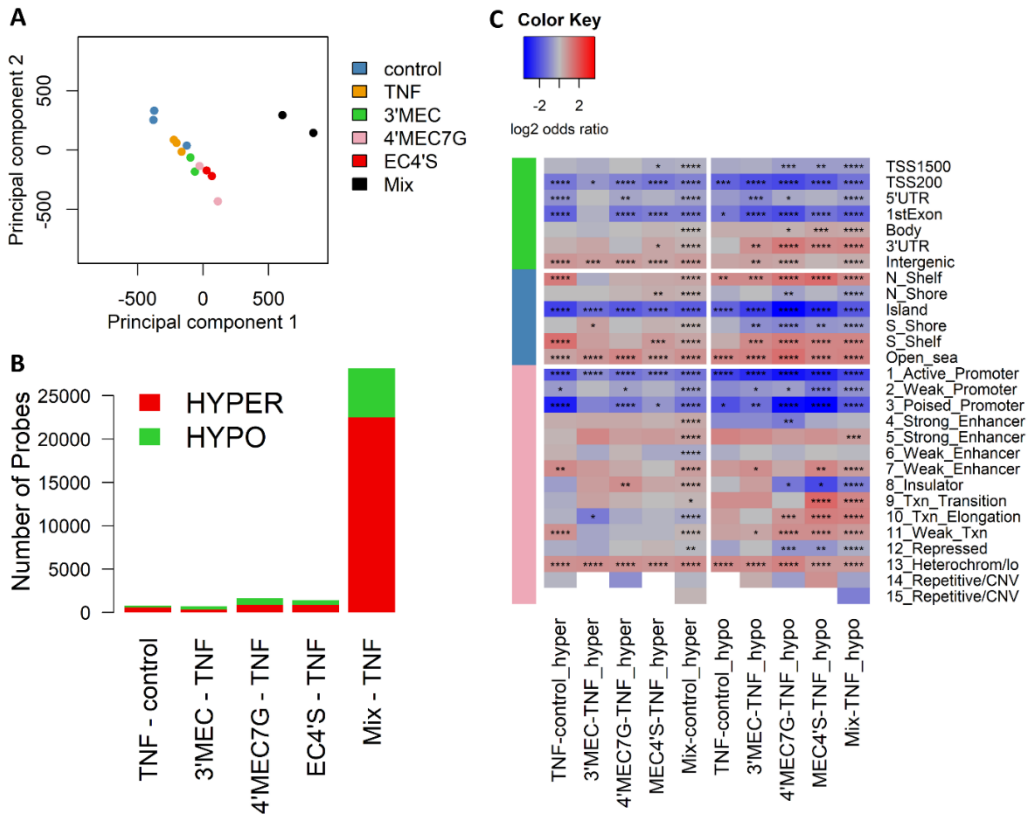


Figure 1: Genome-wide DNA methylation of TNFα-stimulated HUVECs treated with flavanol metabolites. **A)** Principal component analysis. **B)** Number of differentially methylated probes (P < 0.01 & |delta beta| > 0.1). **C)** Enrichment of DMPs in genomic regions. The log2 odds ratios are represented.

The DMPs were mainly located outside promoter sites and CGIs irrespective of the flavanol treatment (**Figure 1C**). DMPs were significantly enriched in CpG poor regions,

intergenic regions, and heterochromatin. In some treatment contrasts, there was a slight enrichment in strong and weak enhancers and in transcriptional chromatin states (9_TxN_Transition, 10_TxN_Elongation and 11_Weak_TxN). No large differences were observed between hyper- and hypo-DMPs.

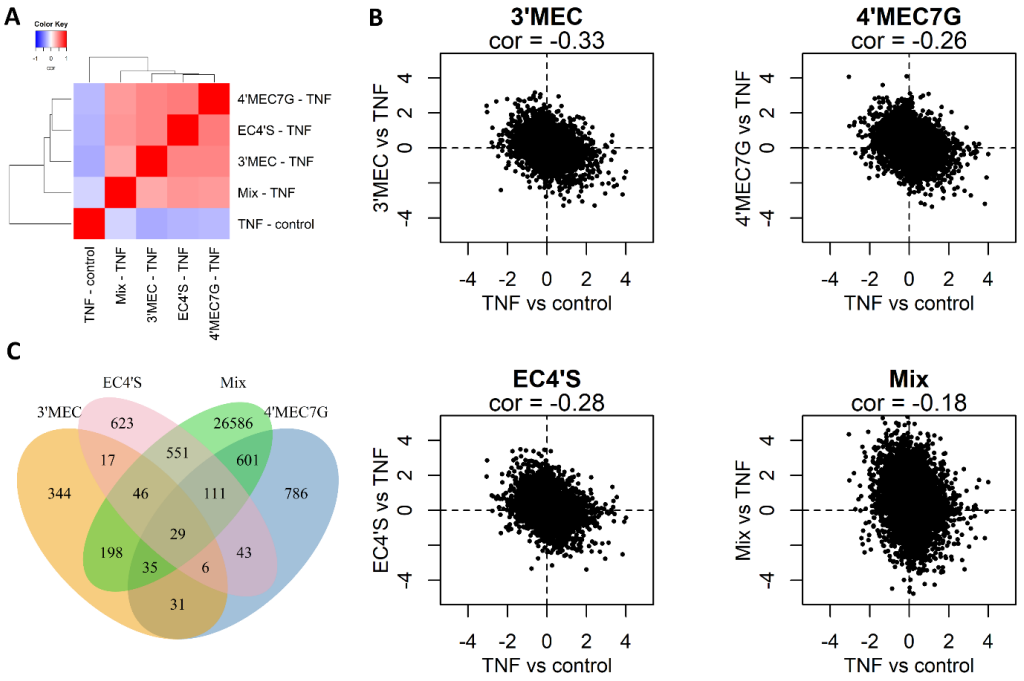


Figure 2: Comparison of DNA methylation effects across different flavanol metabolites. A) Correlation heatmap showing the correlations between the -log10(p-values) across the different treatment contrasts. **B)** Correlation between the -log10(p-values) of the TNF vs control contrast and flavanol metabolite vs TNF contrast. **C)** Venn diagram of the overlapping DMPs across the four flavanol metabolite contrasts.

Flavanol metabolites reverse TNF inflammatory response in endothelial cells

To compare the effect of flavanol metabolites on DNA methylation in HUVECs, we calculated -log10 p-value for every probe, and made this value negative when hypomethylated while hypermethylated probes were kept positive. In this way the most significant hypermethylated CpG probes received the highest positive values, while the most significant hypomethylated CpG probes received the lowest negative values, Next, we performed pairwise correlations of the -log10 p-values between every treatment contrast (**Figure 2A**). We found strong positive correlations among the single and mixture flavanol metabolite contrasts, suggesting that every flavanol metabolite exerts a similar effect on HUVECs DNA methylation profile. In addition, many DMPs

overlapped across the different flavanol treatments (**Figure 2B**). In contrast, there was a negative correlation between flavanol treatment contrast and TNF contrast, suggesting that the effect of the TNF inflammatory response is at least partially reversed by the flavanol metabolites (**Figure 2C**). We also noticed that the correlation between the MIX contrast and the other single flavanol metabolite contrasts was less strong. Similar, the reverse effect of the MIX contrast in respect with the TNF-α response was also limited compared to the single flavanol metabolites.

Flavanol metabolites change endothelial DNA methylation in genes involved in cell adhesion, cytoskeletal organization, renin angiotensin, nitric oxide and axonal guidance signaling pathways

The DMPs were mapped to genes and for every flavanol treatment contrast pathway analysis was performed using IPA. We noticed a strong overlap in significant enriched pathways among the contrasts. To compare the enriched pathways among the different contrasts, we ranked the pathways based on the sum of the -log10 enrichment p-values in each treatment contrast. In this way, the pathways which were enriched in more than one contrast, and are therefore in common across the treatments, were ranked first (**Figure 3**). Among the top ranked pathways, we find interesting pathways including axonal guidance, netrin, ephrin and integrin signaling, actin cytoskeleton signaling, etc., involved in various aspects of atherogenesis via eliciting endothelial dysfunction, monocyte attraction, leukocyte infiltration, monocyte-macrophage retention, platelet hyperreactivity, and neovascularization [37-41]. Previously, we have demonstrated *in vivo* and *in vitro* that epicatechin metabolites mediate their vasculo-protective effects through dynamic regulation of endothelial cell monocyte adhesion and permeability [42]. Furthermore, renin-angiotensin and nitric oxide signaling play critical roles in hypertension and may contribute in cardioprotective blood pressure-lowering effects of dietary flavanols [13, 43-45].

Grape seed flavanol diet intervention triggers strong interindividual variation in blood DNA methylation response.

A double-blind, randomized, placebo-controlled intervention study recently revealed, that the daily consumption of monomeric and oligomeric flavanols (MOF) derived from seeds of grapes (*Vitis vinifera* L.) for 8 weeks accomplish a vascular health benefit in male smokers [17]. Corresponding DNA methylation levels of blood samples were

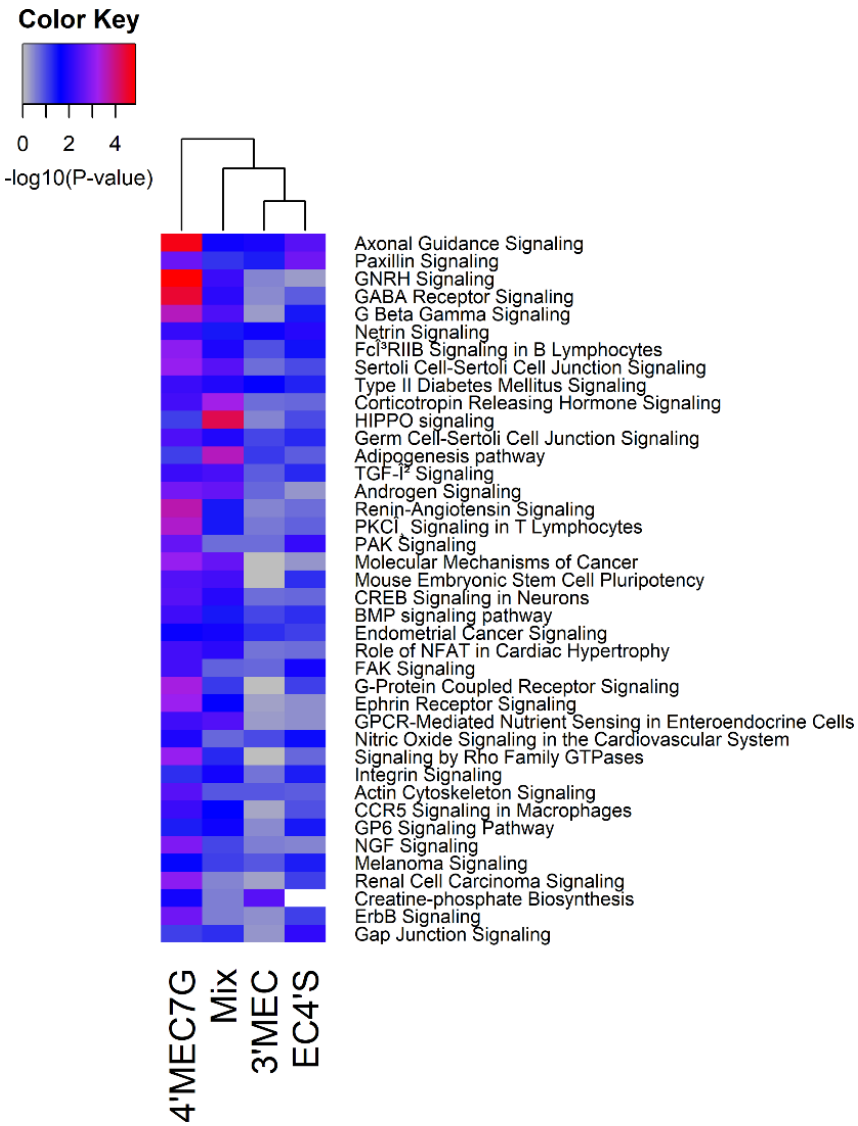


Figure 3: Pathway analysis. Top 40 most significantly enriched IPA canonical pathways across the four flavanol metabolites. The heatmap represents the -log10(p-value) of the enriched pathways. Pathways were ranked by the sum of -log10(p-values).

measured before and after the diet intervention using the 450k Illumina array platform [18]. Clinical characteristics of the volunteers can be found in supplementary table 6. Upon PCA of the DNA methylation profiles obtained in the diet intervention study (data available at GSE54690), the strongest clustering was observed per participant rather than response to diet intervention (**Figure 4A**). In line with previous findings, when we performed linear regression using the limma paired moderated t-test, no common diet responsive CpG probe was found to be differentially methylated (FDRs for all CpG sites were around 1). Similarly, no clear correlation was found of delta beta before and after

diet intervention across the individuals (**Figure 4B**), again showing strong interindividual variation in the epigenetic response to a flavanol enriched diet. Next, we selected probes that changed in methylation by at least 0.1 in beta-value after diet intervention (>10% methylation change) for every individual participant of the study. We noticed that some individuals (SN_110, SN_109 and SN_114) showed a strong epigenetic response with a total number of DMPs between 5,469 and 15,189, while other individuals showed much less difference in methylation changes after the diet intervention (**Figure 4C**). Altogether, we could not identify a common CpG probe which was differentially methylated in every individual, and only a small fraction of the probes was affected in more than one individual (**Supplementary Figure 1**).

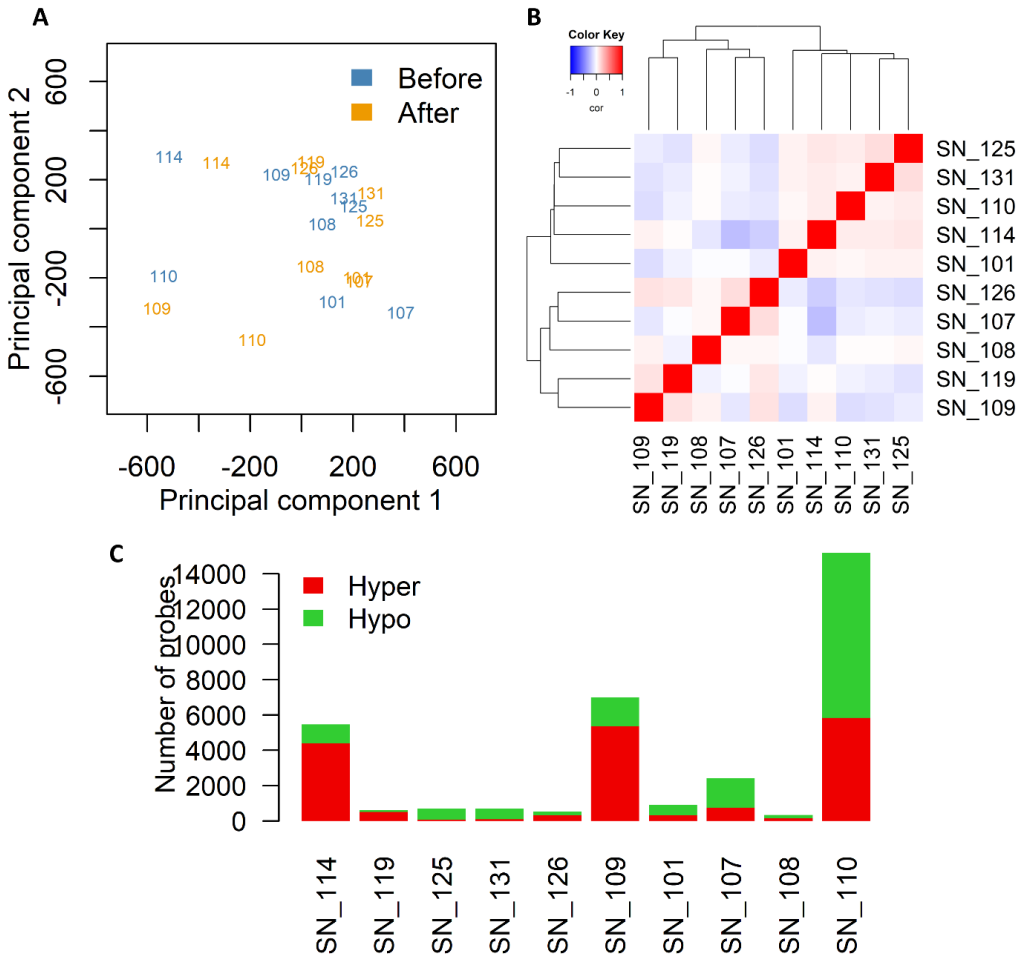


Figure 4: Genome-wide DNA methylation after an 8-week grape seed diet intervention. A) Principal component analysis. Blue labels are the samples before the diet intervention and orange labels are the samples after diet intervention. **B)** Correlation heatmap showing the correlations between the delta betas (methylation differences) across the ten individuals. **C)** Number of DMPs (|delta beta| > 0.1).

Grape seed flavanol diet doesn't increase DNA methylation variability

In another EWAS, differential variable CpG positions were identified which discriminate disease and health conditions [46]. Similarly, we checked whether the grape seed diet had an influence on the methylation variability of specific CpG probes. We therefore looked for CpG probes that had either a lower or higher variability in DNA methylation after the diet intervention as compared to before using the iEVORA algorithm. The iEVORA algorithm makes use of the Bartlett's test to identify differentially variable CpG sites (VMPs). In this way, we identified 336 VMPs (**Figure 5A**). However, upon dividing the VMPs in probes with a higher variability after diet intervention (hyper-VMPs) and probes with a decreased variability after diet intervention (hypo-VMPs), slightly more hypo-VMPs than hyper-VMPs were detected, indicating that the grape seed diet intervention didn't increase the methylation variability (**Figure 5B**). In addition, the Bartlett's test is sensitive to single outliers (which may be due to false positives) and therefore the iEVORA algorithm implements a feature selection step using the differentially methylated t-test p-value. No single CpG probe survived the Bartlett's FDR < 0.05 and t-test p-value < 0.05 selection criteria (**Figure 5A**), indicating that grape seed diet intervention didn't result in a strong change in DNA methylation variability.

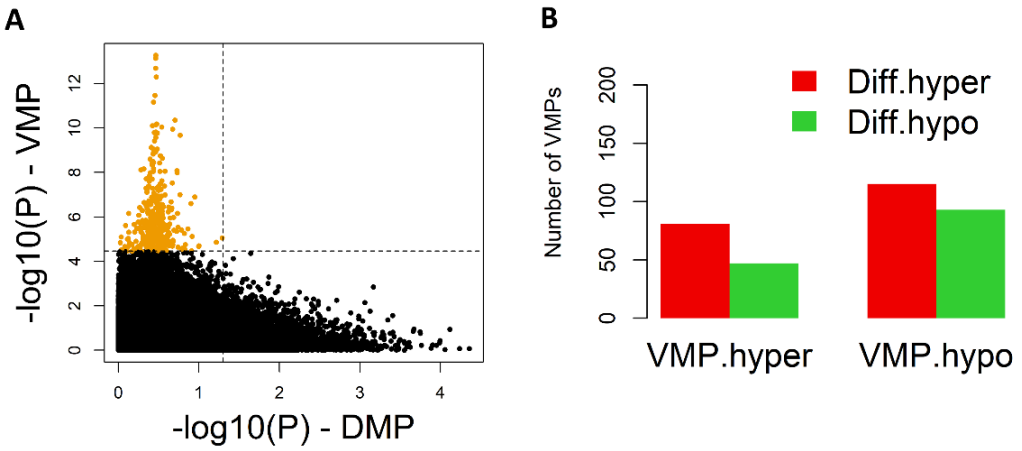


Figure 5: Variable methylated CpG probes. A) Comparison of variable methylated positions (VMPs) and differentially methylated positions (DMPs). Significant VMPs are colored in orange. **B)** Number of VMPs divided in hyper (more variable after diet) and hypo-variable (less variable after diet). Red bars represent the probes which are hypermethylated after the diet intervention and green bars those which were hypomethylated.

Strong interindividual epigenetic variation in response to flavanol diet intervention targets common cell adhesion signaling pathways

For every individual participant we selected the CpG probes that showed a change in methylation of at least 10%, and mapped these probes to genes based on the Illumina

manifest file. Next, we performed IPA pathway analysis for every individual. To compare the enriched pathways among the individuals, we ranked the pathways based on the sum of the -log10 enrichment p-values in each individual. In this way, the pathways which were enriched in more than one individual were ranked first (**Figure 6A**). Among the top ranked pathways, we could find several pathways of interest including leukocyte extravasation signaling, axonal guidance signaling, netrin signaling, thrombin signaling, etc.

Interestingly, leukocyte extravasation signaling, which is known to be influenced by flavanols was also found as one of the top ranked pathways. Leukocyte extravasation is the process of capture, rolling, adhesion and transmigration of leukocytes through the capillary endothelial layer to the place of inflammation. Although this pathways was significantly enriched ($p < 0.05$) in five individuals, little overlap could be observed in the differentially methylated genes involved in this pathway (**Figure 6B**).

Interestingly, 14 of the 40 top enriched canonical pathways were also found to be enriched in the *in vitro* study, including axonal guidance and netrin signaling.

Reciprocal atherosclerosis and flavanol diet specific epigenetic changes driven by shifts in blood cell type composition

Since the cardioprotective role of flavanols is well established, we evaluated whether atherosclerosis and flavanol diet intervention associated DNA methylation profiles show opposite methylation direction. We therefore compared the mean DNA methylation differences after the diet intervention with the DNA methylation differences in atherosclerotic patients of athDMPs.

Interestingly, based on the mean delta betas of all individuals, a negative correlation ($r=-0.41$) was found with the delta betas of the athDMPs (**Figure 7A**). Hypo-athDMPs were on average increased in methylation after the diet intervention, while hyper-athDMPs were mainly decreased in methylation. Furthermore, when calculating the correlation between delta betas in every individual separately, we found more individuals with negative correlation compared as with positive correlations (**Figure 7B and Supplementary Figure 2**). Especially individual SN_114 demonstrated a strong negative correlation. Interestingly, the three samples with the strongest DNA methylation response (SN_114, SN_109 and SN_110) were also the most negatively correlated with the athDMPs.

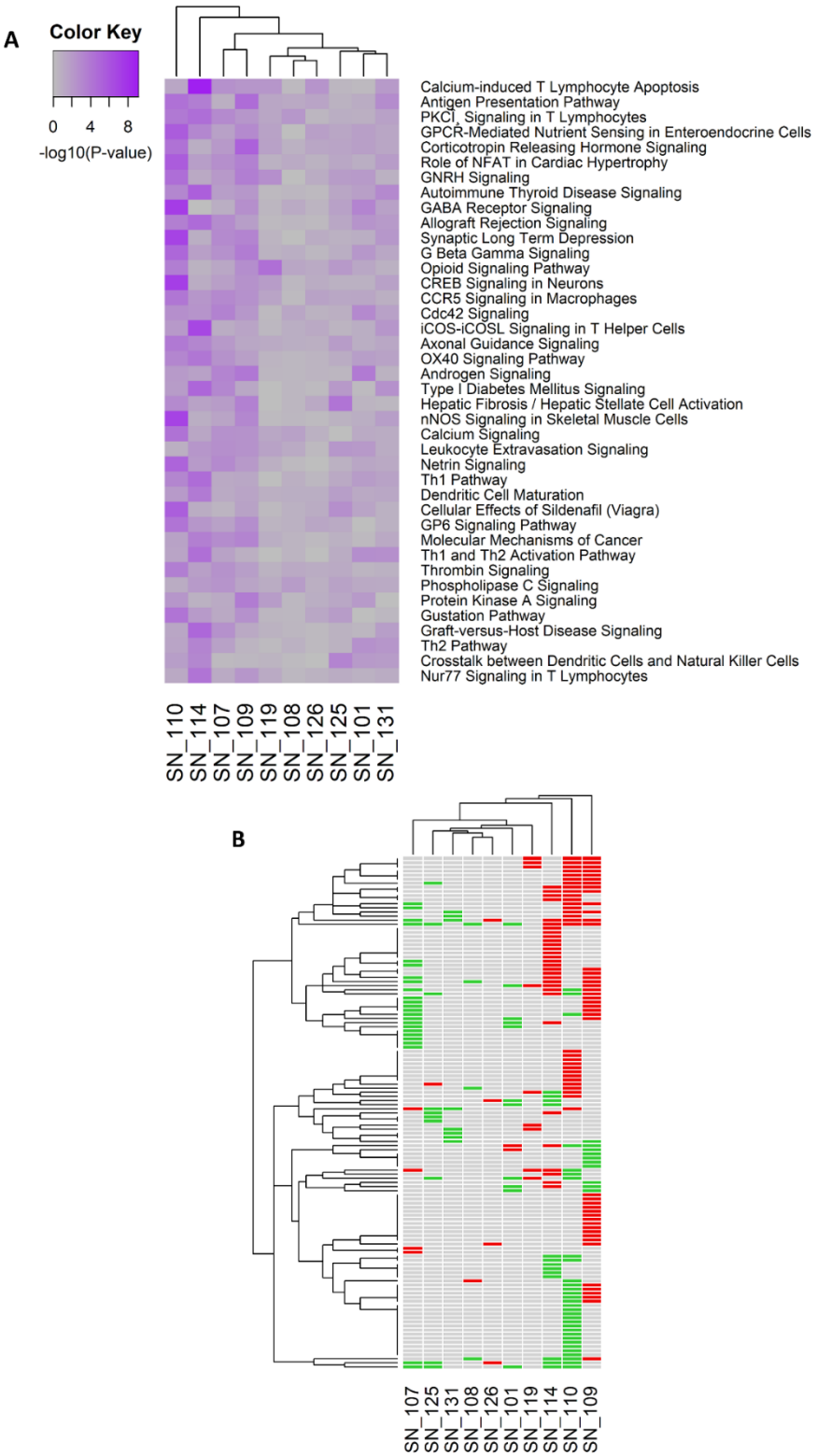

A
Color Key
0
4
8
-log10(P-value)
Calcium-induced T Lymphocyte Apoptosis
Antigen Presentation Pathway
PKCθ Signaling in T Lymphocytes
GPCR-Mediated Nutrient Sensing in Enteroendocrine Cells
Corticotropin Releasing Hormone Signaling
Role of NFAT in Cardiac Hypertrophy
GNRH Signaling
Autoimmune Thyroid Disease Signaling
GABA Receptor Signaling
Allograft Rejection Signaling
Synaptic Long Term Depression
G Beta Gamma Signaling
Opioid Signaling Pathway
CREB Signaling in Neurons
CCR5 Signaling in Macrophages
Cdc42 Signaling
iCOS-iCOSL Signaling in T Helper Cells
Axonal Guidance Signaling
OX40 Signaling Pathway
Androgen Signaling
Type I Diabetes Mellitus Signaling
Hepatic Fibrosis / Hepatic Stellate Cell Activation
nNOS Signaling in Skeletal Muscle Cells
Calcium Signaling
Leukocyte Extravasation Signaling
Netrin Signaling
Th1 Pathway
Dendritic Cell Maturation
Cellular Effects of Sildenafil (Viagra)
GP6 Signaling Pathway
Molecular Mechanisms of Cancer
Th1 and Th2 Activation Pathway
Thrombin Signaling
Phospholipase C Signaling
Protein Kinase A Signaling
Gustation Pathway
Graft-versus-Host Disease Signaling
Th2 Pathway
Crosstalk between Dendritic Cells and Natural Killer Cells
Nur77 Signaling in T Lymphocytes
SN_110
SN_114
SN_107
SN_109
SN_119
SN_108
SN_126
SN_125
SN_101
SN_131
B
SN_107
SN_125
SN_131
SN_108
SN_126
SN_101
SN_119
SN_114
SN_110
SN_109

Figure 6: Pathway analysis of grape-seed diet DNA methylation effects. A) Top 40 most significant enriched IPA canonical pathways. The heatmap represents the -log10(p-value) of the enriched pathways. Pathways were ranked by the sum of -log10(p-values). **B)** Heatmap of the genes in the "Leukocyte Extravasation Signaling" pathway. Genes which were differentially methylated after the diet intervention were colored in red when hypermethylated and green when hypomethylated.

We next used the houseman method to estimate blood cell type composition before and after diet intervention in every individual. As expected, the change in cell type composition after diet intervention was strongly associated with the correlation coefficients between flavanol effect and atherosclerosis in the athDMPs (**Figure 7C**). For example, the change in granulocyte fraction was positively correlated with the correlation coefficients, so that, individual SN_114 with the strongest negative correlation showed the biggest shift in granulocyte fraction, while the samples with positive correlations (SN_107, SN_108 and SN_126) showed as small increase in granulocyte fraction after diet intervention. Similar, the change in CD4+T cell fraction was negatively correlated with the correlation coefficients. As previously shown [47], the relative amount of granulocytes was higher in atherosclerosis, while the relative amount of CD4+T cells was lower in atherosclerosis compared to healthy individuals.

Grape seed DNA methylation effects show weak correlation with vascular health biomarkers

We next checked whether the response of the grape seed flavanol diet on DNA methylation was correlated with vascular function. For this, correlations were performed between the number of DMPs and all the vascular biomarkers which have been monitored in these individuals [17]. No major correlations could be found (**Supplementary Figure 3A**). The strongest positive correlations were found for GPX1 expression values and ED50 of Acetylcholine (Ach.respED50). The most negative correlation was found for endothelin-1 (ET1) levels. A similar analysis was performed using the correlation coefficient with the athDMP instead of the number of DMPs as a measure of grape seed-DNA methylation response (**Supplementary Figure 3B**). As expected, the strongest positive correlation was found for the neutrophil counts and the most negative correlation was found for the lymphocyte counts. Here, GPX1 gene expression levels were found negatively correlated. No association could be found between the grape-seed DNA methylation response and the vascular health index, which is an integrative measure of vascular health combining the results from all the vascular biomarkers measured (**Figure 8A-B**) [17].

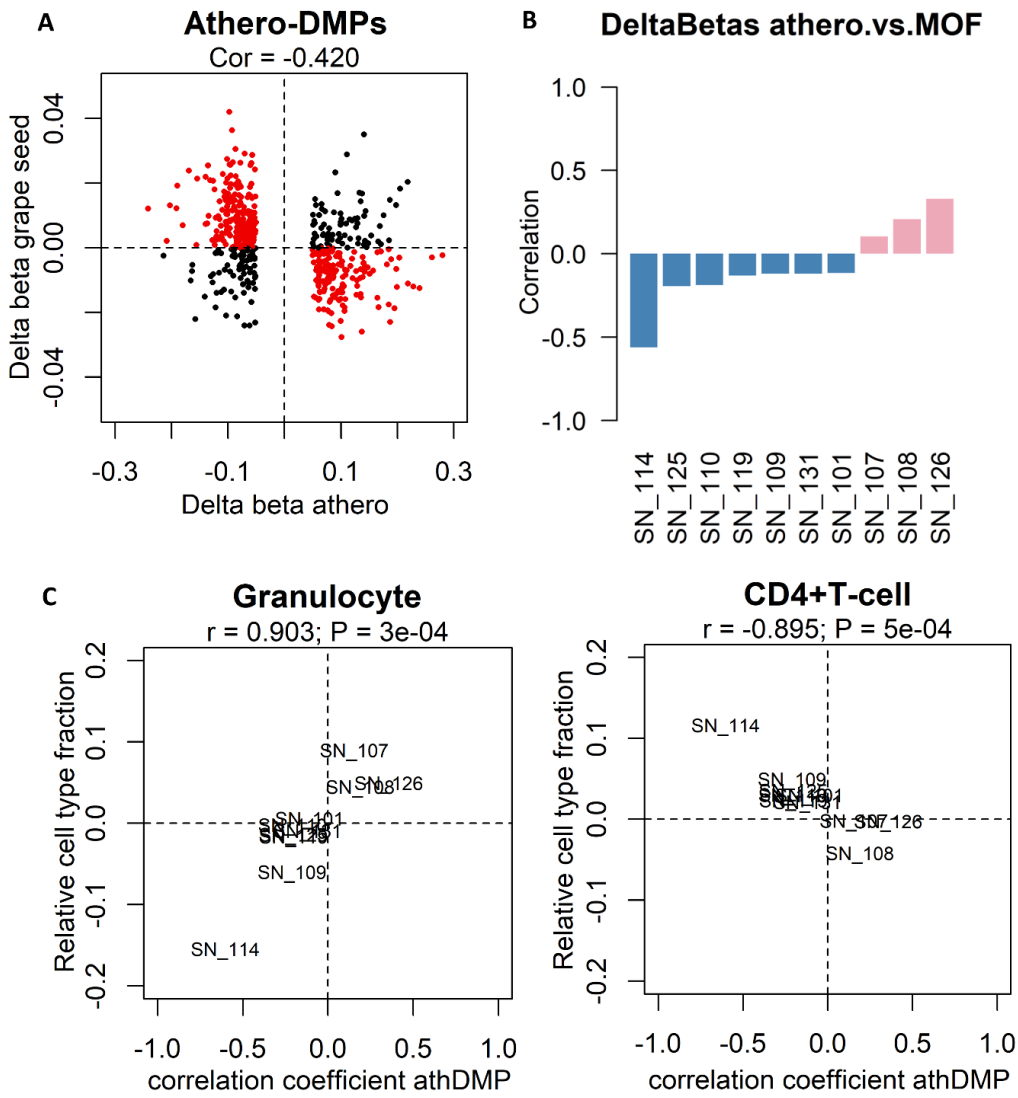


Figure 7: Comparison flavanol diet effect with atherosclerosis DNA methylation effect. A) Comparison of delta betas atherosclerosis of the athero-DMPs (X-axis) and delta betas after grape seed diet intervention (Y-axis). Red CpG probes are those athero-DMPs which were reversed by the diet intervention. **B)** Correlation of the delta betas between atherosclerosis and diet for every individual. **C)** Comparison of the correlation coefficient between atherosclerosis and diet effect (X-axis) and estimated granulocyte (left) and CD4+ T-cell fraction (Y-axis).

Because all study participants were heavy smokers, we also tested whether the flavanol diet could reverse known smoking-induced DNA methylation in blood samples [48]. No clear reversal of the smoking methylation pattern could be observed when comparing average methylation changes of the diet intervention and smoking methylation changes (**Figure 8C**). More specifically, in some individuals a high fraction of smoking-

associated DNA methylation changes was reversed while in other individuals only a small fraction was reversed (**Figure 8D**).

We next checked whether the flavanol diet could impact epigenetic age measured by the epigenetic clock of Horvath [25]. We observed a strong correlation of epigenetic age with chronological age (r = 0.849, P = 2.23e-6). However, no significant change in epigenetic age could be observed after diet intervention (**Figure 8E**).

Discussion

Flavanols are protective against CVDs and improve vascular health, however, the underlying mechanism is not fully elucidated, and it is not yet clear whether the beneficial effects of flavanols are related to epigenetic DNA methylation changes. Our results provide evidence that flavanols can induce DNA methylation changes in endothelial cells *in vitro* and in a diet intervention study *in vivo*, targeting common cell adhesion signaling pathways. Although the flavanol enriched grape seed intervention study reveals no common epigenetic diet responsive signature because of strong inter-individual variation in DNA methylation changes, similar pathways are affected by all individuals. This suggest highly redundant epigenetic modulation of cell adhesion functions.

In HUVECs, flavanol metabolites induced consistent DNA methylation changes. We found a strong overlap of DNA methylation changes across the different flavanol metabolites. Of special note, the mixture of all flavanol metabolites had the strongest effect. The CpG sites that changed after flavanol treatment were rather enriched in intergenic regions and gene bodies, as compared to promoter regions. As a consequence, only a limited number of genes were found to be both differentially expressed and methylated upon comparing genes with DNA methylation changes and gene expression changes, since few promoters showed changes in methylation status. Although additional chromatin dependent promoter regulation cannot be excluded [49], flavanol specific DNA methylation changes may alternatively prime long-term transcriptional responses upon repetitive exposure until reaching a particular DNA methylation threshold. DNA methylation is inherited every cell division and may be seen as an important mediator in memorizing past environmental exposures. Alternatively, it needs to be further investigated whether flavanol specific changes in DNA methylation may have additional roles in alternative splicing or epitranscriptomic events of regulatory lncRNAs [50-53]. Besides, we also found a weak enrichment of DMPs in enhancer regions, especially by the MIX treatment, which could change insulator functions and higher order architecture of TADs to rewire coregulated gene

expression networks [54]. Along the same line, DNA methylation signatures was found to follow preformed chromatin compartmentalization [55].

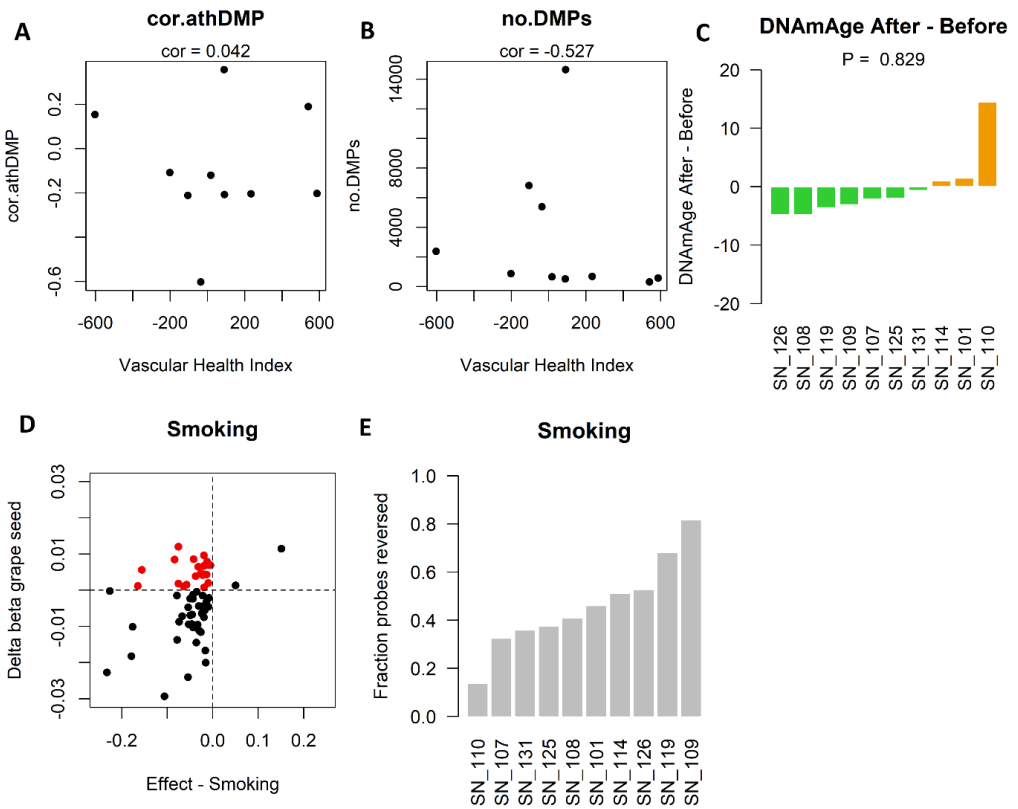


Figure 8: Correlation grape seed diet DNA methylation effect and vascular health markers. A) Comparison vascular health index (X-axis) and the correlation coefficient between atherosclerosis and diet DNA methylation effect (cor.athDMP, Y-axis) in each individual. **B)** Comparison vascular health index (X-axis) and number of DMPs (no.DMPs, Y-axis) in each individual. **C)** Comparison of DNA methylation changes in smokers (X-axis) and after grape seed diet (Y-axis) in known smoking-induced CpG probes. CpG probes reversed by the diet intervention are colored in red. **D)** Fraction of smoking-DMPs reversed by the diet intervention in each individual. E) Difference in DNA methylation age after the diet intervention. Negative values indicate lower epigenetic age after the diet intervention as compared to before the diet intervention.

Of special interest, the intergenic and gene body localized DMPs are enriched for genes involved in cell adhesion related pathways, including axonal guidance signaling, paxillin signaling, netrin signaling, integrin signaling, and actin cytoskeleton signaling. These results are in accordance with previous functional studies showing reduced monocyte-endothelial cell-cell adhesion after flavanol treatment *in vitro* and *in vivo* [42]. The recruitment and migration of monocytes into the arterial wall is one of the key initial processes during the development of an atherosclerotic plaque. Previous studies showed that environmental factors, including LDL cholesterol, homocysteine and

disturbed blood flow may adversely modulate endothelial DNA methylation [56-62]. Interestingly, we further show that the epigenetic impact of an inflammatory stimulus like TNF-α on DNA methylation could partially be suppressed by pretreatment with flavanol metabolites [42].
Furthermore, *in vivo* DNA methylation profiles of an eight-week grape seed diet intervention in ten male smokers were reanalyzed in more depth [18]. The small sample size is a limitation of the study and the data should therefore be seen as pilot study. PCA showed strong interindividual variation in DNA methylation profiles before and after flavanol diet intervention. As such, we failed to detect a consistent epigenetic profile in response to a flavanol rich diet. Moreover, all individuals were heavy smokers with a different smoking history, which may also contribute in high interindividual DNA methylation variability.

However, although we did not identify common flavanol diet specific DNA methylation changes across all individuals, we observed significant gene enrichment in common cell signaling pathways related to cell adhesion and leukocyte extravasation in all participants. This reveals highly redundant epigenetic flexibility of cell adhesion regulation through different gene networks. Epigenetic impact (magnitude, number of DMPs) of flavanol diet also shows strong variation in different individuals. However, the epigenetic impact of flavanols may therefore also depend on history of additional lifestyle factors (nutrition, stress, exercise, smoking behavior, age) which shape the epigenome in a stochastic cumulative fashion [63]. Similarly, monozygotic twin studies reveal age dependent epigenetic drift and epigenome divergence in response to different lifestyles which translates in different disease susceptibilities [64, 65]. Epigenetic drift may also change gene expression levels of many drug metabolizing (ADME genes) and transporter genes, which will affect pharmacokinetics of flavanol consumption [66, 67]. Response to weight-loss interventions have also been associated with DNA methylation and transcriptomic differences [68, 69]. Another study showed that vitamin D response was dependent on the methylation levels of the Cytochrome P450 enzymes CYP2R1 and CYP24A1 [70]. In relation to cancer drug response, MGMT DNA methylation predicts temozolomide response in glioblastoma [71, 72]. Although more research is needed, these examples illustrate the need of DNA methylation markers in personalized nutrition. Considering the stochastic nature of DNA methylation changes, we cannot exclude that a relative short diet intervention (8 weeks) is not sufficient to trigger major changes (>5%) in DNA methylation (our cut-off criterion for differential methylation) as compared to DNA methylation changes following many years (>15 y) of heavy smoking, hypertension and/or aging. Another important parameter is the timing of the intervention. It is known for example that

prenatal and early development are more vulnerable and/or plastic to epigenetic changes than adult stage, to reverse adverse epigenetic marks and gene expression [73].

As such, we found that a flavanol rich diet intervention in adult participants only partially suppresses adverse epigenetic marks associated with atherosclerosis. Furthermore, this could mainly be attributed to decreased granulocyte and increased CD4+T-cell counts. The NLR is a measure of systemic inflammation and is a known prognostic and predictive marker for CVDs. Higher NLR has been associated with increased cardiovascular risk. DNA methylation may be a sensitive method to estimate small shifts in immune cell counts and activation status, and therefore could be a measure for chronic systemic inflammation. Indeed, a methylation-based NLR value was found to predict various cancers. In addition, CRP, another systemic inflammation marker, was shown to be associated with DNA methylation in blood. Interestingly, diet has been shown to have an impact on NLR and chronic inflammation [74]. For example, studies found negative correlations between polyphenol, flavonoid and dark chocolate consumption and low-grade inflammation biomarkers including neutrophil and lymphocyte counts and CRP [75-77]. In rats fed a high-fat diet, grape seed procyanidins could decrease CRP levels [78, 79]. It can therefore be hypothesized that changes in systemic inflammation may also be visible in blood DNA methylation profiles. Whether reduction in inflammation could be measured using DNA methylation should therefore be further investigated in larger study populations.

At the pathophysiological level, we did not detect statistical significant correlations between the reversal of atherosclerosis-associated DNA methylation marks with vascular health improvement or changes in inflammatory markers. Also the number of DMPs in each individual was not correlated with improved vascular health outcome. However, our limited sample size may lack statistical power and limit strong conclusions.

In conclusion, we have demonstrated that flavanols are able to change DNA methylation profiles in endothelial cells *in vitro* and in blood samples *in vivo*, following diet intervention in human smokers. Strong interindividual epigenetic variation in response to flavanol diet intervention further revealed highly redundant effects on cell adhesion networks, beyond direct regulatory effects on gene expression. Finally, an eight week flavanol diet intervention in adult participants, showed limited epigenetic efficacy to suppress adverse epigenetic profiles associated with atherosclerosis or smoking. Longitudinal studies in larger study cohorts are needed to evaluate persistence of observed epigenetic changes and possible causal relations with vascular health parameters.

References

1. Del Rio, D., et al., *Dietary (poly)phenolics in human health: structures, bioavailability, and evidence of protective effects against chronic diseases.* Antioxid Redox Signal, 2013. **18**(14): p. 1818-92.
2. Heiss, C., C.L. Keen, and M. Kelm, *Flavanols and cardiovascular disease prevention.* Eur Heart J, 2010. **31**(21): p. 2583-92.
3. Mangels, D.R. and E.R. Mohler, 3rd, *Catechins as Potential Mediators of Cardiovascular Health.* Arterioscler Thromb Vasc Biol, 2017. **37**(5): p. 757-763.
4. Dower, J.I., et al., *Dietary epicatechin intake and 25-y risk of cardiovascular mortality: the Zutphen Elderly Study.* Am J Clin Nutr, 2016. **104**(1): p. 58-64.
5. Arts, I.C., et al., *Dietary catechins in relation to coronary heart disease death among postmenopausal women.* Epidemiology, 2001. **12**(6): p. 668-75.
6. Cifuentes-Gomez, T., et al., *Factors Affecting the Absorption, Metabolism, and Excretion of Cocoa Flavanols in Humans.* J Agric Food Chem, 2015. **63**(35): p. 7615-23.
7. Schroeter, H., et al., *(-)-Epicatechin mediates beneficial effects of flavanol-rich cocoa on vascular function in humans.* Proc Natl Acad Sci U S A, 2006. **103**(4): p. 1024-9.
8. Desch, S., et al., *Effect of cocoa products on blood pressure: systematic review and meta-analysis.* Am J Hypertens, 2010. **23**(1): p. 97-103.
9. Ried, K., P. Fakler, and N.P. Stocks, *Effect of cocoa on blood pressure.* Cochrane Database Syst Rev, 2017. **4**: p. CD008893.
10. Fisher, N.D., et al., *Flavanol-rich cocoa induces nitric-oxide-dependent vasodilation in healthy humans.* J Hypertens, 2003. **21**(12): p. 2281-6.
11. Rassaf, T., et al., *Vasculoprotective Effects of Dietary Cocoa Flavanols in Patients on Hemodialysis: A Double-Blind, Randomized, Placebo-Controlled Trial.* Clin J Am Soc Nephrol, 2016. **11**(1): p. 108-18.
12. Tokede, O.A., J.M. Gaziano, and L. Djousse, *Effects of cocoa products/dark chocolate on serum lipids: a meta-analysis.* Eur J Clin Nutr, 2011. **65**(8): p. 879-86.
13. Jimenez, R., J. Duarte, and F. Perez-Vizcaino, *Epicatechin: endothelial function and blood pressure.* J Agric Food Chem, 2012. **60**(36): p. 8823-30.
14. Sansone, R., et al., *Cocoa flavanol intake improves endothelial function and Framingham Risk Score in healthy men and women: a randomised, controlled, double-masked trial: the Flaviola Health Study.* Br J Nutr, 2015. **114**(8): p. 1246-55.
15. Rodriguez-Mateos, A., et al., *Assessing the respective contributions of dietary flavanol monomers and procyanidins in mediating cardiovascular effects in humans: randomized, controlled, double-masked intervention trial.* Am J Clin Nutr, 2018. **108**(6): p. 1229-1237.
16. Ramirez-Sanchez, I., et al., *(-)-Epicatechin induced reversal of endothelial cell aging and improved vascular function: underlying mechanisms.* Food Funct, 2018. **9**(9): p. 4802-4813.
17. Weseler, A.R., et al., *Pleiotropic benefit of monomeric and oligomeric flavanols on vascular health--a randomized controlled clinical pilot study.* PLoS One, 2011. **6**(12): p. e28460.
18. Milenkovic, D., et al., *Dietary flavanols modulate the transcription of genes associated with cardiovascular pathology without changes in their DNA methylation state.* PLoS One, 2014. **9**(4): p. e95527.
19. Claude, S., et al., *Flavanol metabolites reduce monocyte adhesion to endothelial cells through modulation of expression of genes via p38-MAPK and p65-Nf-kB pathways.* Mol Nutr Food Res, 2014. **58**(5): p. 1016-27.
20. Szarc vel Szic, K., et al., *From inflammaging to healthy aging by dietary lifestyle choices: is epigenetics the key to personalized nutrition?* Clin Epigenetics, 2015. **7**: p. 33.
21. Fang, M.Z., et al., *Tea polyphenol (-)-epigallocatechin-3-gallate inhibits DNA methyltransferase and reactivates methylation-silenced genes in cancer cell lines.* Cancer Res, 2003. **63**(22): p. 7563-70.

22. Nandakumar, V., M. Vaid, and S.K. Katiyar, *(-)-Epigallocatechin-3-gallate reactivates silenced tumor suppressor genes, Cip1/p21 and p16INK4a, by reducing DNA methylation and increasing histones acetylation in human skin cancer cells.* Carcinogenesis, 2011. **32**(4): p. 537-44.
23. Khan, M.A., et al., *(-)-Epigallocatechin-3-gallate reverses the expression of various tumor-suppressor genes by inhibiting DNA methyltransferases and histone deacetylases in human cervical cancer cells.* Oncol Rep, 2015. **33**(4): p. 1976-84.
24. Sharma, U. and O.J. Rando, *Metabolic Inputs into the Epigenome.* Cell Metab, 2017. **25**(3): p. 544-558.
25. Horvath, S., *DNA methylation age of human tissues and cell types.* Genome Biol, 2013. **14**(10): p. R115.
26. Horvath, S. and K. Raj, *DNA methylation-based biomarkers and the epigenetic clock theory of ageing.* Nat Rev Genet, 2018. **19**(6): p. 371-384.
27. Chen, B.H., et al., *DNA methylation-based measures of biological age: meta-analysis predicting time to death.* Aging (Albany NY), 2016. **8**(9): p. 1844-1865.
28. Declerck, K. and W. Vanden Berghe, *Back to the future: Epigenetic clock plasticity towards healthy aging.* Mech Ageing Dev, 2018. **174**: p. 18-29.
29. Quach, A., et al., *Epigenetic clock analysis of diet, exercise, education, and lifestyle factors.* Aging (Albany NY), 2017. **9**(2): p. 419-446.
30. Szarc Vel Szic, K., et al., *Epigenetic silencing of triple negative breast cancer hallmarks by Withaferin A.* Oncotarget, 2017. **8**(25): p. 40434-40453.
31. Assenov, Y., et al., *Comprehensive analysis of DNA methylation data with RnBeads.* Nat Methods, 2014. **11**(11): p. 1138-1140.
32. Teschendorff, A.E., et al., *A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data.* Bioinformatics, 2013. **29**(2): p. 189-96.
33. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and microarray studies.* Nucleic Acids Res, 2015. **43**(7): p. e47.
34. Teschendorff, A.E., et al., *DNA methylation outliers in normal breast tissue identify field defects that are enriched in cancer.* Nat Commun, 2016. **7**: p. 10478.
35. Houseman, E.A., et al., *DNA methylation arrays as surrogate measures of cell mixture distribution.* BMC Bioinformatics, 2012. **13**: p. 86.
36. Teschendorff, A.E., et al., *A comparison of reference-based algorithms for correcting cell-type heterogeneity in Epigenome-Wide Association Studies.* BMC Bioinformatics, 2017. **18**(1): p. 105.
37. Poitz, D.M., et al., *EphrinB2/EphA4-mediated activation of endothelial cells increases monocyte adhesion.* Mol Immunol, 2015. **68**(2 Pt C): p. 648-56.
38. Lin, Z., et al., *Netrin-1 prevents the attachment of monocytes to endothelial cells via an anti-inflammatory effect.* Mol Immunol, 2018. **103**: p. 166-172.
39. Okamoto, T. and K. Suzuki, *The Role of Gap Junction-Mediated Endothelial Cell-Cell Interaction in the Crosstalk between Inflammation and Blood Coagulation.* Int J Mol Sci, 2017. **18**(11).
40. Hu, S., et al., *Vascular Semaphorin 7A Upregulation by Disturbed Flow Promotes Atherosclerosis Through Endothelial beta1 Integrin.* Arterioscler Thromb Vasc Biol, 2018. **38**(2): p. 335-343.
41. Hu, S. and L. Zhu, *Semaphorins and Their Receptors: From Axonal Guidance to Atherosclerosis.* Front Physiol, 2018. **9**: p. 1236.
42. Milenkovic, D., et al., *A systems biology network analysis of nutri(epi)genomic changes in endothelial cells exposed to epicatechin metabolites.* Sci Rep, 2018. **8**(1): p. 15487.
43. Litterio, M.C., et al., *Blood pressure-lowering effect of dietary (-)-epicatechin administration in L-NAME-treated rats is associated with restored nitric oxide levels.* Free Radic Biol Med, 2012. **53**(10): p. 1894-902.
44. Flammer, A.J., et al., *Cardiovascular effects of flavanol-rich chocolate in patients with heart failure.* Eur Heart J, 2012. **33**(17): p. 2172-80.
45. Nehme, A. and K. Zibara, *Cellular distribution and interaction between extended renin-angiotensin-aldosterone system pathways in atheroma.* Atherosclerosis, 2017. **263**: p. 334-342.
46. Paul, D.S., et al., *Increased DNA methylation variability in type 1 diabetes across three immune effector cell types.* Nat Commun, 2016. **7**: p. 13555.

47. Istas, G., et al., *Identification of differentially methylated BRCA1 and CRISP2 DNA regions as blood surrogate markers for cardiovascular disease.* Sci Rep, 2017. **7**(1): p. 5120.
48. Gao, X., et al., *DNA methylation changes of whole blood cells in response to active smoking exposure in adults: a systematic review of DNA methylation studies.* Clin Epigenetics, 2015. **7**: p. 113.
49. Declerck, K., et al., *Epigenetic control of cardiovascular health by nutritional polyphenols involves multiple chromatin-modifying writer-reader-eraser proteins.* Curr Top Med Chem, 2016. **16**(7): p. 788-806.
50. Ji, P., et al., *N6-Methyladenosine in RNA and DNA: An Epitranscriptomic and Epigenetic Player Implicated in Determination of Stem Cell Fate.* Stem Cells Int, 2018. **2018**: p. 3256524.
51. Maunakea, A.K., et al., *Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition.* Cell Res, 2013. **23**(11): p. 1256-69.
52. Linker, S.M., et al., *Combined single-cell profiling of expression and DNA methylation reveals splicing regulation and heterogeneity.* Genome Biol, 2019. **20**(1): p. 30.
53. Singh, S., et al., *Intragenic DNA methylation and BORIS-mediated cancer-specific splicing contribute to the Warburg effect.* Proc Natl Acad Sci U S A, 2017. **114**(43): p. 11440-11445.
54. Wiehle, L., et al., *DNA (de)methylation in embryonic stem cells controls CTCF-dependent chromatin boundaries.* Genome Res, 2019. **29**(5): p. 750-761.
55. Nothjunge, S., et al., *DNA methylation signatures follow preformed chromatin compartments in cardiac myocytes.* Nat Commun, 2017. **8**(1): p. 1667.
56. Yu, X., W. Ling, and M. Mi, *Relationship of impairment induced by intracellular S-adenosylhomocysteine accumulation with DNA methylation in human umbilical vein endothelial cells treated with 3-deazaadenosine.* Int J Exp Pathol, 2009. **90**(6): p. 638-48.
57. Yang, T.C., et al., *Malondialdehyde mediates oxidized LDL-induced coronary toxicity through the Akt-FGF2 pathway via DNA methylation.* J Biomed Sci, 2014. **21**: p. 11.
58. Ma, S.C., et al., *Homocysteineinduced oxidative stress through TLR4/NFkappaB/DNMT1mediated LOX1 DNA methylation in endothelial cells.* Mol Med Rep, 2017. **16**(6): p. 9181-9188.
59. Kumar, A., et al., *Histone and DNA methylation-mediated epigenetic downregulation of endothelial Kruppel-like factor 2 by low-density lipoprotein cholesterol.* Arterioscler Thromb Vasc Biol, 2013. **33**(8): p. 1936-42.
60. Jiang, Y.Z., et al., *Hemodynamic disturbed flow induces differential DNA methylation of endothelial Kruppel-Like Factor 4 promoter in vitro and in vivo.* Circ Res, 2014. **115**(1): p. 32-43.
61. Castro, R., et al., *Intracellular S-adenosylhomocysteine increased levels are associated with DNA hypomethylation in HUVEC.* J Mol Med (Berl), 2005. **83**(10): p. 831-6.
62. Jiang, Y.Z., et al., *Arterial endothelial methylome: differential DNA methylation in athero-susceptible disturbed flow regions in vivo.* BMC Genomics, 2015. **16**: p. 506.
63. Feinberg, A.P. and R.A. Irizarry, *Evolution in health and medicine Sackler colloquium: Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease.* Proc Natl Acad Sci U S A, 2010. **107 Suppl 1**: p. 1757-64.
64. Fraga, M.F., et al., *Epigenetic differences arise during the lifetime of monozygotic twins.* Proc Natl Acad Sci U S A, 2005. **102**(30): p. 10604-9.
65. Zheng, S.C., M. Widschwendter, and A.E. Teschendorff, *Epigenetic drift, epigenetic clocks and cancer risk.* Epigenomics, 2016. **8**(5): p. 705-19.
66. Tang, X. and S. Chen, *Epigenetic Regulation of Cytochrome P450 Enzymes and Clinical Implication*. Curr Drug Metab, 2015. **16**(2): p. 86-96.
67. Habano, W., et al., *Analysis of DNA methylation landscape reveals the roles of DNA methylation in the regulation of drug metabolizing enzymes.* Clin Epigenetics, 2015. **7**: p. 105.
68. Bouchard, L., et al., *Differential epigenomic and transcriptomic responses in subcutaneous adipose tissue between low and high responders to caloric restriction.* Am J Clin Nutr, 2010. **91**(2): p. 309-20.
69. Moleres, A., 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, 2013. **27**(6): p. 2504-12.

70. Zhou, Y., et al., *DNA methylation levels of CYP2R1 and CYP24A1 predict vitamin D response variation.* J Steroid Biochem Mol Biol, 2014. **144 Pt A**: p. 207-14.
71. Hegi, M.E., et al., *MGMT gene silencing and benefit from temozolomide in glioblastoma.* N Engl J Med, 2005. **352**(10): p. 997-1003.
72. Esteller, M., et al., *Inactivation of the DNA-repair gene MGMT and the clinical response of gliomas to alkylating agents.* N Engl J Med, 2000. **343**(19): p. 1350-4.
73. Bianco-Miotto, T., et al., *Epigenetics and DOHaD: from basics to birth and beyond.* J Dev Orig Health Dis, 2017. **8**(5): p. 513-519.
74. Wang, R., P.J. Chen, and W.H. Chen, *Diet and exercise improve neutrophil to lymphocyte ratio in overweight adolescents.* Int J Sports Med, 2011. **32**(12): p. 982-6.
75. Pounis, G., et al., *Polyphenol intake is associated with low-grade inflammation, using a novel data analysis from the Moli-sani study.* Thromb Haemost, 2016. **115**(2): p. 344-52.
76. Chun, O.K., et al., *Serum C-reactive protein concentrations are inversely associated with dietary flavonoid intake in U.S. adults.* J Nutr, 2008. **138**(4): p. 753-60.
77. di Giuseppe, R., et al., *Regular consumption of dark chocolate is associated with low serum concentrations of C-reactive protein in a healthy Italian population.* J Nutr, 2008. **138**(10): p. 1939-45.
78. Terra, X., et al., *Modulatory effect of grape-seed procyanidins on local and systemic inflammation in diet-induced obesity rats.* J Nutr Biochem, 2011. **22**(4): p. 380-7.
79. Terra, X., et al., *Grape-seed procyanidins prevent low-grade inflammation by modulating cytokine expression in rats fed a high-fat diet.* J Nutr Biochem, 2009. **20**(3): p. 210-8.

Supplementary information

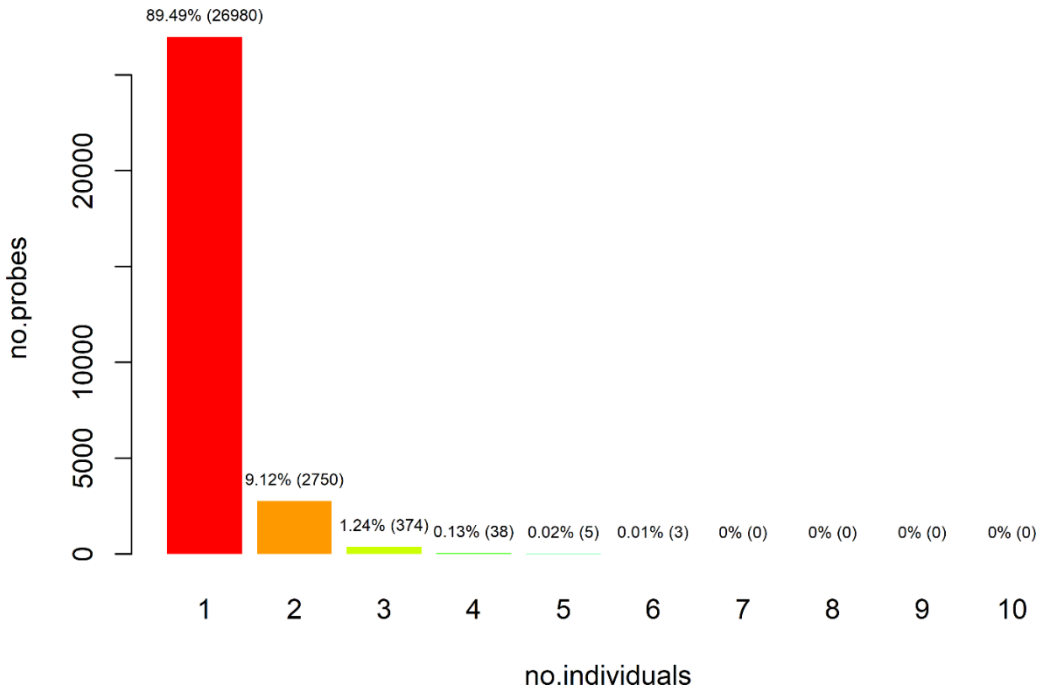
Supplementary tables 1 to 5 can be found in following dropbox folder: https://www.dropbox.com/sh/d9lq9mipqkoezx3/AAC9RCXgWS7P6DS71ybJqWNJa?dl=0

Supplementary Table 1-5: DMPs of contrasts 3'MEC vs TNF, 4'MEC7G vs TNF, EC4'S vs TNF, Mix vs TNF and TNF vs control.

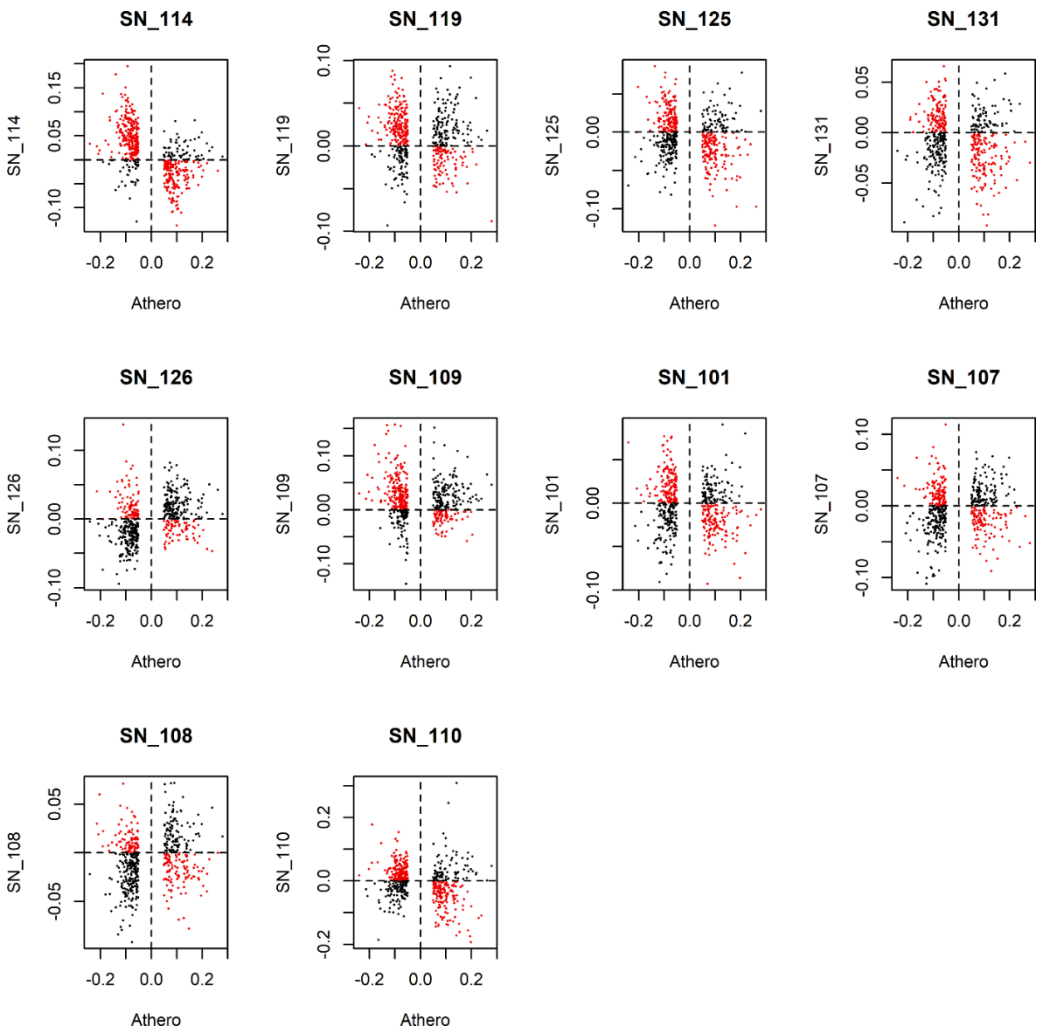
Supplementary Table 6: Anthropometric and clinical characteristics of the male volunteers.

	Men (n = 13)	**Clinical reference values**
Age , y	48 (30-58)	NA
BMI, kg/m	25 (18-28)	<30
Years of smoking	28 (14-42)	NA
Cigarettes/day	16 (10-25)	NA
Pack years	19 (7-52)	NA
SBP, mmHg	115 (103-124)	<140
DBP, mmHg	74 (60-85)	<90
Heart rate, beats/min	55 (50-66)	60-100
tChol, mmol/L	5.7 (4.0-7.1)	≤5.0
LDL, mmol/L	3.9 (2.5-5.5)	≤2.5
HDL, mmol/L	1.2 (0.9-4.0)	0.9-1.7
TG, mmol/L	1.2 (0.9-4.0)	<20
CRP, mg/L	1.7 (0.6-5.6)	<10
Fibrinogen, g/L	3.4 (2.4-4.8)	2.0-4.0

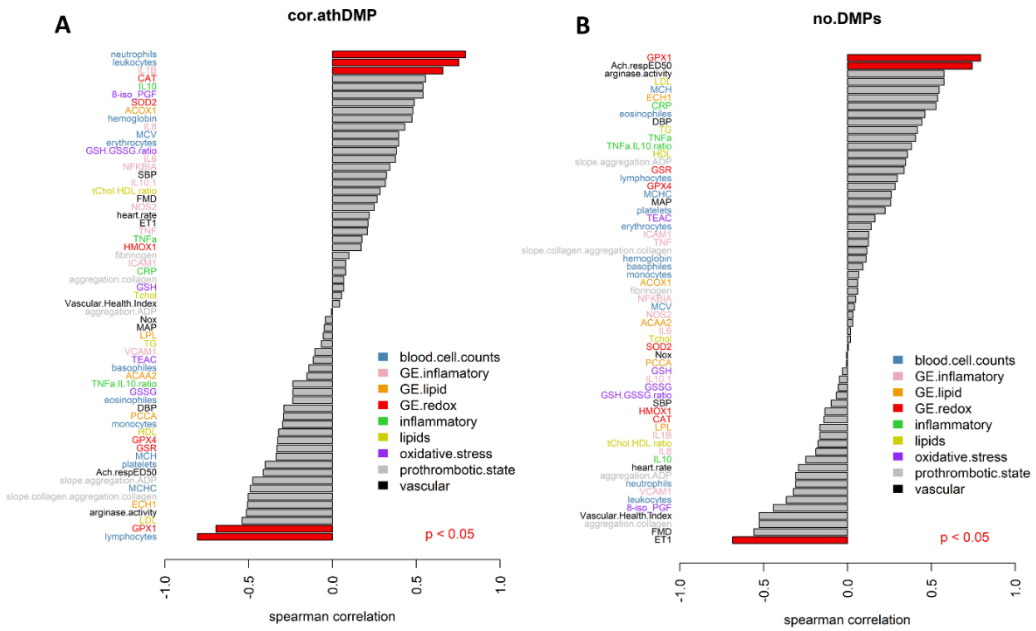
Values are median (range).
Clinical reference values are Dutch reference values "clinical chemistry" (www.flk.cvz.nl) and from the Dutch General Practitioner Guidelines (https://www.nhg.org/standaarden/volledig/cardiovasculair-risicomanagement).



Supplementary Figure 1: Number of probes differentially methylated in one or multiple individuals.



Supplementary Figure 2: Correlation between diet effect and atherosclerosis effect on DNA methylation for each individual.



Supplementary Figure 3: Correlation between diet effect on DNA methylation and vascular health markers. **A)** diet effect was measured as the athDMP correlation coefficient. **B)** diet effect was measured as the number of DMPs. Vascular health markers were grouped in different categories (blood.cell.counts, GE.inflammatory, etc).

8

Standardized medicinal *Echinacea purpurea* tincture enforces monocyte innate immune response through priming of interferon and chemotaxis pathways

Ken Declerck[1], Claudina Perez Novo[1] , Guy Van Camp[2], Andreas Suter[3] and Wim Vanden Berghe[1]

[1] Laboratory of Protein Chemistry, Proteomics and Epigenetic Signaling (PPES), Department of Biomedical Sciences, University of Antwerp (UA), Belgium
[2] Center of Medical Genetics, Department of Biomedical Sciences, University of Antwerp (UA), Belgium
[3] A. Vogel Bioforce AG, Roggwil, Switzerland

Under revision in Frontiers in Immunology.

Contributions: Ken Declerck performed the experiments, analyzed the data, interpreted the results and wrote the manuscript.

Abstract | Herbal medicinal *Echinacea purpurea* tinctures are widely used today to prevent and treat respiratory tract infections. Although their immunomodulatory effects have already been demonstrated by several studies, the molecular mechanisms involved remain poorly characterized. Integration of transcriptome, kinome and DNA-methylome profiles of THP1 monocyte cells treated with a quality-controlled preparation of *Echinacea purpurea* extract (Echinaforce®), revealed activation of an innate antiviral immune gene response. More particularly, Echinaforce® treatment of THP1 cells triggers interferon signaling pathways and gene expression changes related to innate immunity. Phosphopeptide based kinome activity profiling and pharmacological inhibitor experiments with filgotinib further demonstrate crucial involvement of JAK1 kinase activation in IFN signaling and immunity related gene activation by Echinacea treatment. Finally, Echinaforce® treatment promotes predominantly DNA hypermethylation at intergenic CpG and DNA repeat elements (LINE, SINE, LTR) and silences transcription of flanking endogenous retroviral sequences (ERVs), as part of an evolutionary conserved (epi)genomic protective response against viral infections. Altogether, Echinaforce® treatment of THP1 cells activates an innate immune response by priming JAK1 dependent interferon signaling and gene expression, DNA repeat hypermethylation and ERV silencing, which might contribute in protection against respiratory infections.

Introduction

Distinct species of the plant genus *Echinacea* have traditionally been used in North America against infectious diseases and wounds [1, 2]. Currently, a wide variety of Echinacea preparations are used world-wide to improve the immune system, and for preventing and treating common colds and influenza infections. Of all Echinacea species, *Echinacea purpurea* (purple coneflower) is the most popular used variety in Western countries.

Different *Echinacea purpurea* extracts (different species, plant parts, manufacturing) or derived compounds showed antioxidant, antibacterial, antifungal, antiviral and mosquitocidal activities in cell culture experiments [3], although absolute comparisons between studies with different preparations remain difficult [4, 5]. A meta-analysis of multiple intervention studies revealed a significant reduction of common cold in the *Echinacea* treated as compared to the placebo control groups [4]. Meta-analysis of six clinical trials showed a reduced risk for recurrence of respiratory tract infections and complications by Echinacea [6].

Complex immunomodulatory actions of Echinacea have been described including both pro- and anti-inflammatory effects [2, 3, 7]. The compounds that contribute to these

activities are alkylamides, glycoproteins, polysaccharides and caffeic acid derivates that may act independently or in synergy [1, 3, 8, 9]. Therefore, different Echinacea preparations with variable constituents may lead to different immune outcomes [5]. For example, a polysaccharide-rich root extract increased the expression of MHC II, CD86 and CD54 markers on mice dendritic cells, while an alkylamide-rich leaf extract suppressed the expression of these surface markers [10].

In this study, we evaluated Echinaforce®, a commercially registered herbal medicinal tincture of *Echinacea purpurea* plant (A.Vogel Bioforce, Switzerland) in several European countries like Switzerland, Austria, UK, Spain, the Netherlands, Denmark, Finland, Sweden, Slovenia as well as Canada, containing 5% root extract and 95% herb extract following extraction with 65% ethanol V/V. Echinaforce® has been shown to possess immune modulatory, anti-inflammatory, anti-bacterial, anti-viral and anti-parasitic activity [11-22]. Clinical efficacy could be shown with different batches in acute treatment [23] or for prevention [24] of respiratory tract infections. A 4-month clinical study on the safety and efficacy of Echinaforce® to prevent common cold showed significant more cold episodes and higher duration in the placebo group as compared to the Echinaforce® treatment group [24]. Recurring infections were lower in the Echinaforce® treatment group. Moreover, no differences between placebo and Echinaforce® group were reported in relation to health risk and safety [24].

Despite the promising immune potentiating properties of Echinaforce®, the responsible molecular targets have only partially been identified. For example, multiple studies demonstrate that *Echinacea* alkylamides exerts their action partially through the cannabinoid receptor 2 (CB2) [8, 20, 25]. Alkylamides are structurally similar to endocannabinoids, and bind CB2 with a higher affinity compared to endogenous cannabinoids [25]. This action of alkylamides on CB2 was the mechanism behind the upregulation of TNF-α in primary monocytes after Echinaforce® treatment [20]. In addition, the cAMP, p38/MAPK and JNK signaling pathways, as well as NF-κB and ATF2/CREB1 TFs were found to be involved in the Echinacea-induced TNF-α expression [20]. Furthermore, a recent study showed that *ex vivo* blood stimulation with LPS after an 8-day oral Echinacea administration resulted in the induction of anti-inflammatory cytokines (IFN-γ, IL-8, IL-10 and MCP-1) only in subjects with low basal levels of these cytokines [11]. Interestingly, individuals with high basal levels of these cytokines didn't show any further induction. The same was seen in subjects with high stress levels or high susceptibility to cold infections, suggesting that Echinaforce® mainly (or only) enhances immune responses in immunocompromised subjects [11].

To further clarify mode of action of standardized Echinacea purpura extract, we applied a systems biology approach by combining genome-wide gene expression, DNA

methylation and kinome activity profiles of THP1 monocyte cells following Echinaforce® treatment.

Materials and methods

Test item

Echinaforce® (batch nr. 040070, A. Vogel Bioforce AG, Roggwil, Switzerland) is a standardized preparation obtained by ethanol extraction of freshly harvested *Echinacea purpurea* herb and roots (95:5). Echinaforce® is marketed as a registered medicinal product and produced under conditions of Good Manufacturing Practice GMP. Thus, a consistent quality for each produced batch is mandatory and equal to the requirements for an allopathic remedy. The composition of marker compounds like alkylamides (i.e. those compounds known to characterize this species of Echinacea) was described previously [3, 7, 26]. In contrast to pressed juice extracts, Echinaforce® extract does not contain polysaccharides which are known to stimulate the immune system unspecifically [27-30]. The alcohol concentration of Echinaforce® tincture extract was 65% v/v and solvent controls have been included in all experimental in vitro experiments to rule out aspecific effects. In addition, the preparation was free of detectable endotoxin as determined by means of a commercial assay kit, Lonza Walkersville Inc., MD, lower limit of detection 0.1 unit/ml.

Cell lines and treatments

THP1 cells were grown in RPMI-1640 medium supplemented with glutamine, 10 % heat inactivated Fetal Bovine Serum, 50 IU/mL Penicillin, 50 µg/mL Streptomycin, 10 mM HEPES and 0.05 mM β-mercaptoethanol. Cells were treated with 1% Echinaforce® tincture versus ethanol solvent control. Each treatment condition consisted of six biological replicates.

Genome-wide gene expression analysis

Sample preparation and microarray processing

THP1 cells were treated for 48h with 1% Echinaforce® or ethanol solvent control. RNA was isolated using the RNeasy mini kit (Qiagen) according to manufacturer's instructions. RNA concentration and purity was measured using the nanodrop 1000. The quality of each RNA sample was checked using the Experion of Bio-Rad. 500 ng of total RNA was amplified using the Illumina TotalPrep RNA Amplification kit (Life

Technologies, Carlsbad, CA, USA). Briefly, RNA was reverse transcribed using T7 oligo(dT) primers, after which biotinylated cRNA was synthesized through an *in vitro* transcription reaction. 750 ng of amplified cRNA was hybridized to a HumanHT12 beadchip array (Illumina, San Diego, CA, USA). The beadchip was incubated for 18 hours at 58 °C in a hybridization oven under continuous rocking. After several consecutive washing steps, bead intensities were read on an Illumina iScan.

Microarray data preprocessing and analysis

Raw gene expression intensities were preprocessed using the beadarray R package [31]. Intensities were quantile normalized and $\log_2$ transformed. Probes with a P-detection value higher than 0.05 in at least six samples were removed. Also, probes annotated as "bad" and "no match" (as described in [32]) were not kept for further analysis. Differentially gene expression was performed using the limma R package [33]. P-values were corrected for multiple testing using the method of Benjamini and Hochberg. Probes with a $\log_2$ fold change higher than 0.4 and an adjusted p-value of 0.05 were defined as significant and kept for further analysis. The probes were annotated with gene information using the illuminaHumanv4.db annotation dataset [34]. The gene IDs of the significant Illumina expression probes were uploaded into the IPA software to find enriched biological pathways, diseases and networks. Raw and normalized array data were uploaded to the Gene Expression Omnibus (GEO) database and have accession number: GSE117904.

Quantitative PCR validation

THP1 cells were treated with 1% Echinaforce® or Solvent for 3-6-12-24-48h in three independent biological experiments. The effect of JAK1 inhibition was determined by treating the cells with 1 µM JAK1 inhibitor Filgotinib (GLPG0634, Selleckchem) for 30 min before adding Echinaforce®. RNA was isolated using the RNeasy mini kit (Qiagen) according to manufacturer's instructions. 750 ng RNA was reverse transcribed into cDNA using oligo dT (Invitrogen), M-MLV reverse transcriptase (Promega), 2.5 mM dNTPs and RNaseOUT (Invitrogen). Samples were incubated on 42°C for 60 min and 75°C for 15 min. For the hERV genes, cDNA synthesis was performed using random primers (Invitrogen) and incubation of the samples at 37°C for 60 min and 75°C for 15 min. qPCR was performed using the GoTaq qPCR Master Mix (Promega) on a StepOnePlus Real-Time PCR machine (Applied Biosystems). Following primers were used: MX1 forward primer (FP) 5'-GTTTCCGAAGTGGACATCGCA-3', MX1 reverse primer (RP) 5'-CTGCACAGGTTGTTCTCAGC-3', IFITM1 FP 5'-CCAAGGTCCACCGTGATTAAC-3', IFITM1 RP 5'-ACCAGTTCAAGAAGAGGGTGTT-3', STAT1 FP 5'-CCATCCTTTGGTACAACATGC-3', STAT1 RP 5'-TGCACATGGTGGAGTCAGG-3', IL8 FP 5'-GCTCTCTTGGCAGCCTTCCTGA-3', IL8 RP 5'-ACAATAATTTCTGTGTTGGCGC-3', CXCL10 FP

5’-GAAAGCAGTTAGCAAGGAAAGGT-3’, CXLC10 RP 5’-GACATATACTCCATGTAGGGAAGTGA-3’, ACTB FP 5’-CTGGAACGGTGAAGGTGACA-3’, and ACTB RP 5’- AAGGGACTTCCTGTAACAATGCA-3’. Primer sequences for hERVs were derived from [35]. Each sample was ran in triplicate and median Ct-values in each replicate group was selected. Ct-values were normalized using ACTB housekeeping gene. ddCt-values or log fold changes (logFC) were calculated using the solvent control as reference sample. Paired t-test was used to determine the significance of the differences between Echinaforce® and solvent expression levels.

Kinase activity profiling

Sample preparation

THP1 cells were treated with 1% Echinaforce® or ethanol solvent control for 15 min. Cell lysates were prepared according to manufacturer’s instructions. In short, cells were washed twice with cold 1X PBS and lysed with lysis buffer (1:100 dilution of Halt Phosphatase Inhibitor Cocktail and Halt Protease Inhibitor Cocktail EDTA free in M-PER Mammalian Extraction Buffer (ThermoFisher Scientific™, Rockford, USA) at a ratio of 100 µl buffer per $1x10^6$ cells. Lysates were then incubated on ice for 15 min and centrifuged for 15 minutes at 16000 x g at 4°C. Protein concentration was quantified using the Pierce BCA Protein Assay Kit (ThermoFisher Scientific™, Rockford, USA).

Serine/threonine kinases (STK) and tyrosine kinase (PTK) pamgene assay and data analysis

Kinase activity profiling was performed PamChip® preprocessing and kinase activity profiling was performed according to manufacturer’s instructions (PamGene International BV, ’s-Hertogenbosch, The Netherlands). The first part of the protocol consisted in the blocking of the arrays with 2% BSA followed by several washing steps. Then 0.5 µg for STK and 5 µg for PTK assays together with the correspondent reaction mixes (purchased from the Pamgene) were loaded onto the arrays and incubated in the microarray system PamStation ® 12 instrument (PamGene International, Den Bosch, The Netherlands). In this step, the ATP contained in the mix leads to the activation of the kinases in the lysate which will result in the phosphorylation of the peptides on the array. Peptide phosphorylation intensities are than detected with the primary STK antibody mix and FITC-labeled antibody for STK assay and with the FITC-labelled PTK antibody (PTK assay). Images are then taken by the CCD camera in the PamStation®12 and processed by the Bionavigator software. Peptide intensities data were $\log_2$ transformed and differences in phosphorylation between Echinaforce® treated and

control cultures were determined by using an univariate student t-test analysis corrected for multiple testing using the Benjamini and Hochberg method [36].
To identify potentially activated or inhibited kinases we used the STK or PTK Upstream Kinase analysis PamApp from the Bionavigator Software. The analysis is based on "*in silico* predictions" for the upstream kinases of phosphorylation sites in the human proteome that are retrieved from the phosphoNET database [37]. In short, a prediction algorithm is derived from known interactions between kinases and phosphorylation sites. The prediction algorithm is then used to predict the strength of undocumented interactions. The Bionavigator application uses PhosphoNet database to map putative kinases upstream of the phospho-peptides (a kinase can have multiple possible phosphosites, and a single site can be phosphorylated by different kinases). For each set of peptides mapped to a specific kinase, a "difference statistics" is calculated (=normalized kinase statistics) using following formula: $\tau = \frac{1}{n}\sum_{i=1}^{n} \frac{\bar{p}_{i1} - \bar{p}_{i2}}{\sqrt{s_{i1}^2 + s_{i2}^2}}$ with $\bar{p}_{ij}$ and $\bar{s}_{ij}$ as the sample mean and variance of the intensity of peptide i in group j, respectively. n are the number of peptides linked with a specific kinase. A positive kinase statistic means that the kinase is activated, while a negative statistic means the kinase is inactivated compared to the control group. The kinases are subsequently ranked based on a specificity and significance score which are calculated using permutation of the peptides and samples, respectively. Following formula is used: $Q = -\log_{10}(\max\left(\frac{m}{M}, \frac{1}{M}\right))$, where m is the number of times out of M permutations that $|\tau_p| > |\tau|$, where τ_p is the value of the difference statistic obtained after permutation of the samples or peptides. The significance score represents the magnitude of the change represented by the normalized kinase statistic. The specificity score represents the specificity of the of normalized kinase statistic in terms of the set of peptides used for the corresponding kinase. The higher the score the less likely it is that the observed normalized kinase statistics could have been obtained using a random set of peptides from the data set. The sum of the significance and specificity score is used to rank the kinases [38].

Genome-wide DNA methylation analysis

Sample preparation

THP1 cells were treated for 48h with 1% Echinaforce® or ethanol solvent control. DNA treated with Echinaforce® was isolated using the DNeasy Blood & Tissue kit (Qiagen) according to manufacturer's instructions. DNA concentration and purity was measured using the nanodrop 1000. 1 µg of DNA was used for bisulfite conversion using the EZ DNA methylation Kit of Zymo Research according to manufacturer's instructions.

Successful bisulfite conversion was checked using a methylation-specific PCR in a region of the SALL3 gene (see [39] for primer sequences).

EPIC DNA methylation array

The Infinium HumanMethylationEPIC BeadChip array (Illumina, San Diego, CA, USA) was used to measure genome-wide DNA methylation. 4 µL of bisulfite-converted DNA from each sample was amplified, fragmented, precipitated, resuspended and subsequently hybridized onto the BeadChips. After overnight incubation of the BeadChips, unhybridized fragments were washed away, while hybridized fragments were extended using fluorescent nucleotide bases. Finally, the BeadChips were scanned using the Illumina iScan system to obtain raw methylation intensities of each probe.

EPIC DNA methylation data preprocessing and analysis

The R package RnBeads was used to preprocess the Illumina 450K methylation data [40]. CpG-probes were filtered before normalization based on following criteria: probes containing a SNP within 3 bp of the analyzed CpG site, bad quality probes based on an iterative greedycut algorithm with a detection p-value threshold of 0.01, and probes with missing values in at least one sample. After filtering these CpG-probes, methylation values were within-array normalized using the beta mixture quantile dilation (BMIQ) method [41]. Another filtering step was performed after normalization based on following criteria: probes measuring methylation not at CpG sites (CC, CAG, CAH, …) and probes on sex chromosomes.

The methylation beta-values were transformed to M-values ($M = \log_2(\beta/(1-\beta))$) prior to further analyses. The moderated t-test incorporated in the limma R package [33] was used to calculate the statistics and p-values of the methylation differences between Echinaforce®- and solvent-treated samples. Significant differentially methylated probes (DMPs) were selected based on a FDR < 0.1 and a difference in beta-value of at least 0.05. The DMPs were annotated with gene information using the IlluminaHumanMethylationEPICmanifest R package [42]. Further gene information was retrieved from the UCSC genome browser (human hg19). Enrichment of genomic regions was calculated using the Fisher's exact test. Pathway analysis of the genes harboring a DMP was performed using the Ingenuity Pathway Analysis (IPA) software. Raw and normalized array data were uploaded to the Gene Expression Omnibus (GEO) database and have accession number: GSE117904.

Results and discussion

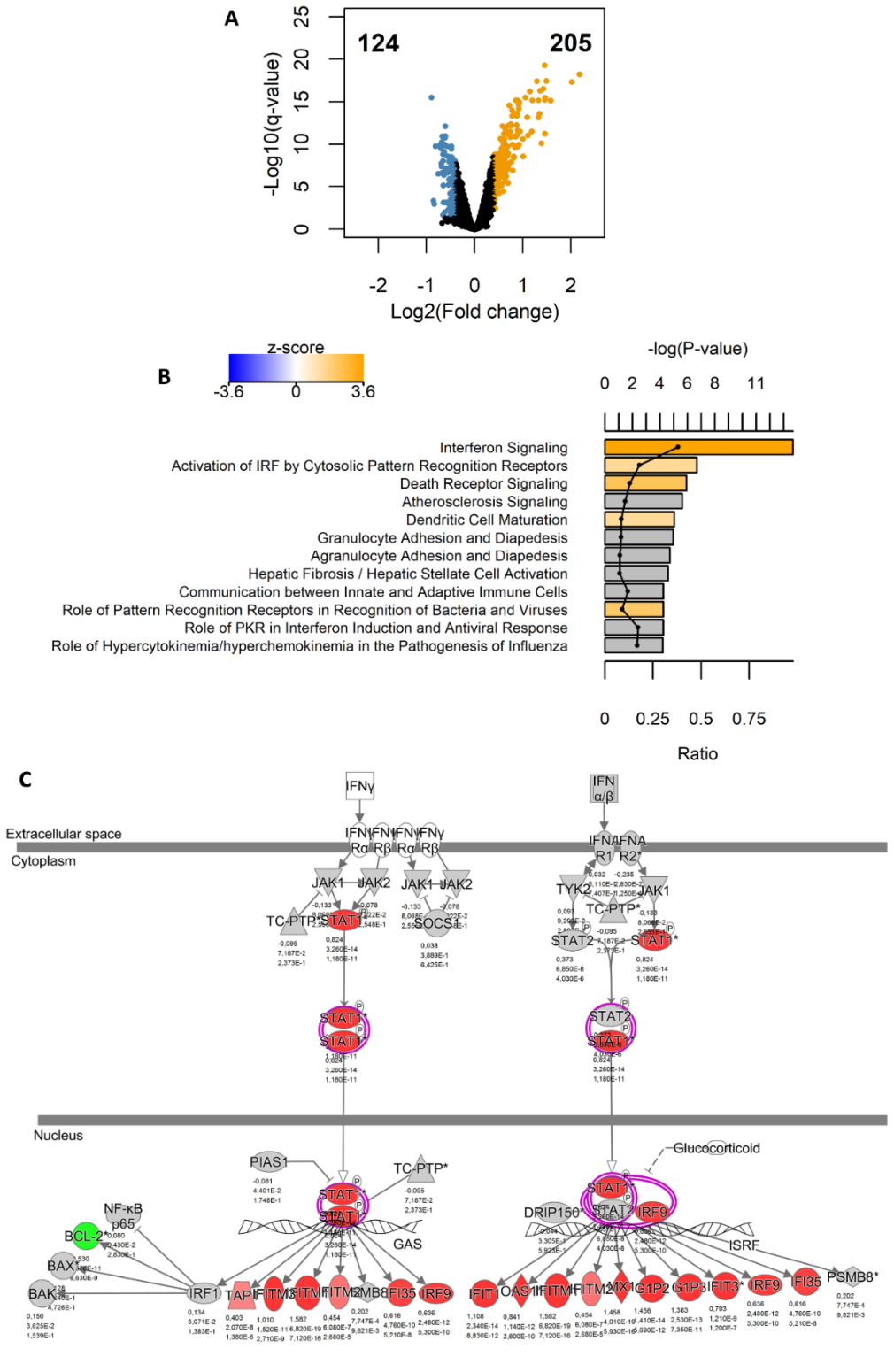
Echinaforce® treatment activates an interferon and chemokine innate immune signaling gene response

Widespread gene expression changes in monocyte THP1 cells were detected upon 48h 1% Echinaforce® treatment. Based on significance criteria of FDR < 0.05 and absolute log_2 fold change > 0.4, Echinaforce® induced modest upregulation of 205 expression probes (173 genes) while 124 probes (99 genes) were downregulated compared with the ethanol treated solvent controls (**Figure 1A and Supplementary table 1**). In contrast to pharmacological drugs (for example glucocorticoids) with can trigger drastic expression changes of specific genes, many phytochemicals typically achieve significant biological responses through mild transcriptional changes of redundant genes converging on the same pathway [43-45]. Genes differentially expressed by Echinaforce® treatment were enriched for IPA canonical pathways related to innate immune responses, including 'interferon signaling', 'activation of IRF by cytosolic pattern recognition receptors', 'dendritic cell maturation', 'granulocyte and agranulocyte adhesion and diapedesis', 'Role of Pattern Recognition Receptors in Recognition of Bacteria and Viruses', etc. (**Figure 1B and Supplementary table 2**). Interestingly, most of these pathways were predicted to be activated, as can be seen from the highly positive activation z-scores.

Especially, the interferon (IFN) signaling pathway was strongly enriched (high significance) and being activated (**Figure 1C**). As a consequence, viral infection and replication was predicted to be inhibited (activation z-score < -2), while cellular movement, migration, recruitment and chemotaxis was predicted to be activated (activation z-scores > 2) (**Figure 1D**). Both STAT1 and STAT2 gene expression show moderate induction, triggering downstream expression of both type I (IFN-α/β) and type II IFN-regulated genes involved in antiviral immune responses.

Also antiviral gatekeeper and interferon inducible proteins MX1 and IFITM1 (CD225) show transcriptional activation [46-48]. Similarly, IL8 and CXCL10 chemokines involved in innate immunity are transcriptionally activated [49].

In line with our results, anti-viral effects against influenza infection and activation of IFN pathways have also been reported *in vivo* following Echinaforce® tincture treatment [11, 15, 21]. In accordance with the canonical pathways, the differentially expressed genes were also enriched in inflammation related diseases and biological functions, including 'infectious diseases', 'immunological diseases', 'cellular movement', 'inflammatory response', 'immune cell trafficking', 'antimicrobial response', etc. (**Supplementary table 3**). In this respect, besides interferon-stimulated


A
124
205
-Log10(q-value)
Log2(Fold change)
B
z-score
-3.6
0
3.6
-log(P-value)
Interferon Signaling
Activation of IRF by Cytosolic Pattern Recognition Receptors
Death Receptor Signaling
Atherosclerosis Signaling
Dendritic Cell Maturation
Granulocyte Adhesion and Diapedesis
Agranulocyte Adhesion and Diapedesis
Hepatic Fibrosis / Hepatic Stellate Cell Activation
Communication between Innate and Adaptive Immune Cells
Role of Pattern Recognition Receptors in Recognition of Bacteria and Viruses
Role of PKR in Interferon Induction and Antiviral Response
Role of Hypercytokinemia/hyperchemokinemia in the Pathogenesis of Influenza
Ratio
C
Extracellular space
Cytoplasm
Nucleus
IFNγ
IFN α/β
JAK1
JAK2
TYK2
TC-PTP
TC-PTP*
STAT1
STAT2
SOCS
PIAS1
Glucocorticoid
DRIP150*
IRF9
GAS
ISRE
NF-κB
p65
BCL-2
BAX
BAK
IRF1
TAP1
IFITM3
IFITM1
IFITM2
MB8
FI35
IRF9
IFIT1
OAS1
IFITM1
IFITM2
MX1
G1P2
G1P3
IFIT3*
IRF9
FI35
PSMB8*

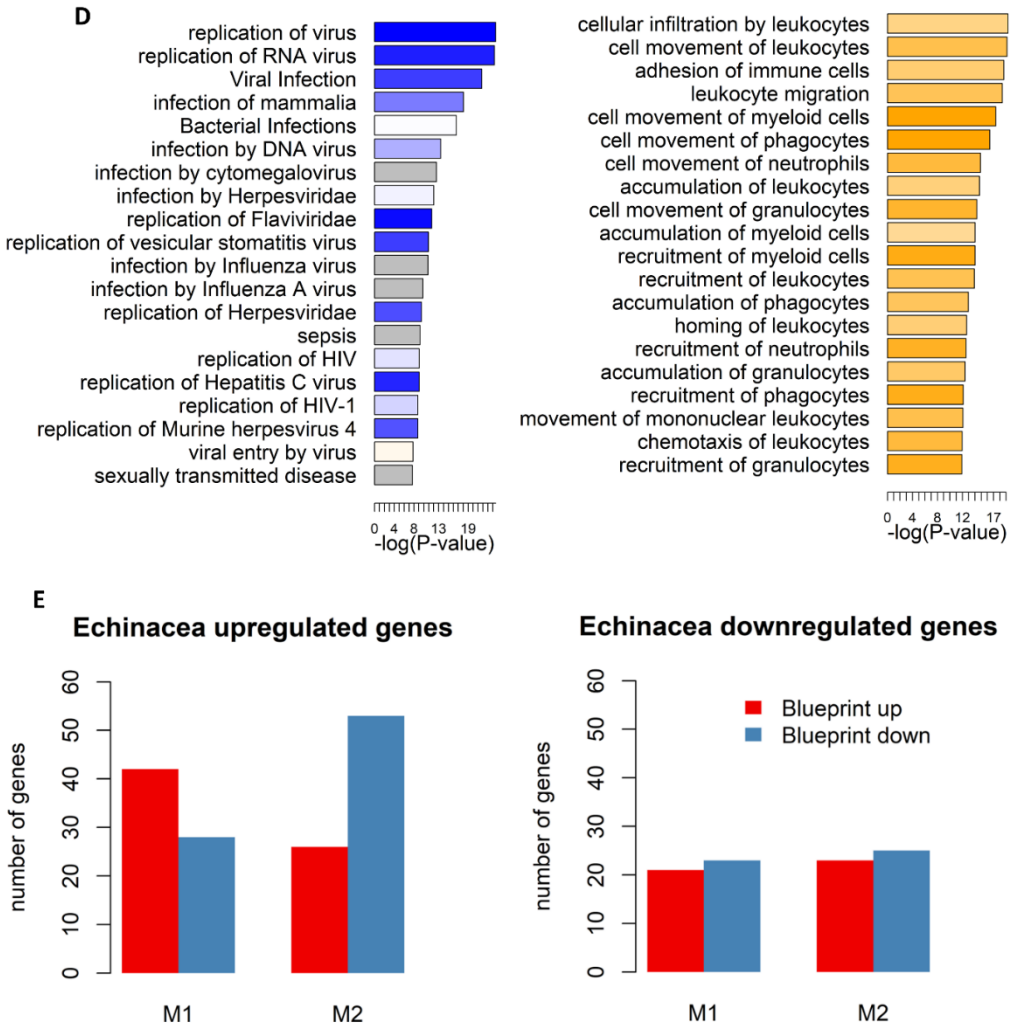


Figure 1: Echinaforce® alters expression of genes involved in the interferon and NF-κB signaling pathway. **A)** Volcano plot showing the upregulated genes (orange color, number of probes: 205), and downregulated genes (blue color, number of probes: 124) upon treatment of THP1 cells for 48h with Echinaforce® tincture (1%) . **B)** Top enriched IPA canonical pathways. Bars are colored by activation z-score. **C)** IPA interferon signaling pathway with Echinaforce®-induced up- and downregulated genes colored in red and green, respectively. **D)** Top enriched IPA infectious diseases and IPA immune trafficking disease and bio function. Bars are colored by activation z-score. **E)** Barplots showing number of Echinaforce®-regulated-genes up- or downregulated in BLUEPRINT M1 and M2 macrophages compared to monocytes.

genes, also a lot of chemokines were upregulated, which promote the recruitment and adhesion of immune cells.

Since IFN signaling plays a key role in monocyte/macrophage polarization [50], we next compared our results with public available gene expression data of CD14+ CD16- monocytes, inflammatory macrophages (M1) and alternative macrophages (M2)

obtained from the BLUEPRINT consortium [51]. Echinaforce®-responsive genes were correlated with gene expression profiles of monocyte/macrophage subtypes by pairwise t-tests. In general, Echinaforce®-upregulated genes were found to be highly expressed in M1 as compared to M2 macrophages. Along the same line, Echinaforce®-downregulated genes were also weakly expressed in M1 macrophages. Of the 205 Echinaforce®-upregulated genes, 172 were also part of the BLUEPRINT dataset, of which 42 genes were also upregulated (Bonferroni p-value < 0.05) in M1 macrophages as compared to monocytes (**Figure 1E**). A smaller number of genes (n=26) were found to be downregulated in M1 macrophages. In contrast, in M2 macrophages the Echinaforce®-upregulated genes were most often downregulated (53 downregulated genes compared to 28 upregulated genes). For the Echinaforce®-downregulated genes, there was no strong difference between up- and downregulated genes in M1 and M2 macrophages. Altogether, this suggests that Echinaforce®-induced gene expression is more closely related to an inflammatory activated macrophage state (M1) then an alternative activated macrophage state (M2) (**Figure 1E**).

To validate microarray results and to evaluate the time dependent transcriptional dynamics of different genes involved in IFN and chemotaxis innate immune signaling, we performed qPCR validation of STAT1, MX1, IFITM1, CXCL8 and CXCL10 transcription levels in THP1 monocytes following different exposure times of Echinacea tincture (3-6-12-24-48h). Gene induction of STAT1 and the interferon-stimulated genes MX1 and IFITM1 could clearly be confirmed, with maximal gene expression observed after 48h treatment (**Figure 2**). Furthermore, persistent increased expression levels of the chemokines IL8 (CXCL8) and CXCL10 were observed, which peaked after 3h with a gradual decrease in gene expression at later time points (**Figure 2**). Along the same line, in murine dendritic cells, Echinacea extract stimulated cell mobility and chemotaxis, and altered expression of cell adhesion and motility genes [52]. Reciprocally, different studies showed the ability of Echinaforce® to reverse the chemokine induction of virus-infected cells [12, 53-55]. While Echinacea induces cytokine and chemokine expression in uninfected cells, their induction is suppressed upon virus infection or LPS stimulation [30, 53-55]. Similarly, Echinaforce® increased the transcription of TNF-α in human monocytes, but reduced the LPS-stimulated TNF-α protein production [20]. Although multiple studies suggest that this stimulatory effect may be the result of bacterial-derived LPS and lipoproteins [27-30], our Echinaforce® tincture contains no polysaccharides and was shown previously to be free of endotoxins, and still exerts an immune activating effect [13, 20]. Overall, our results suggest that Echinaforce® treatment activates an interferon and chemokine innate immune signaling gene response.

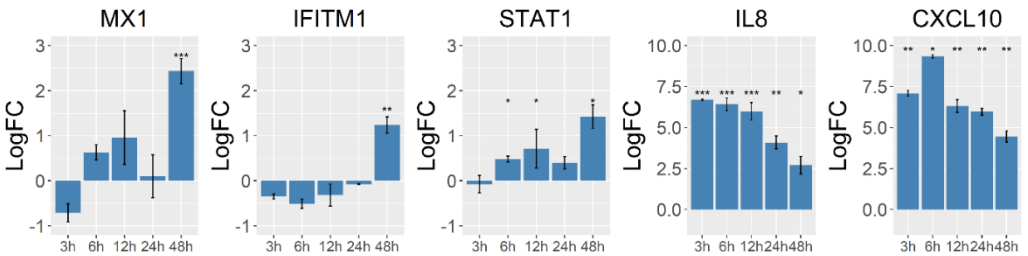


Figure 2: qPCR validation Echinacea-induced upregulated genes at different time points. The mean logFC values +- SD compared to the solvent control are represented in bars. * P ≤ 0.05, ** P ≤ 0.01, *** P ≤ 0.001 and **** P ≤ 0.0001.

Echinaforce® activates IFN, NFKB and MAPK kinases

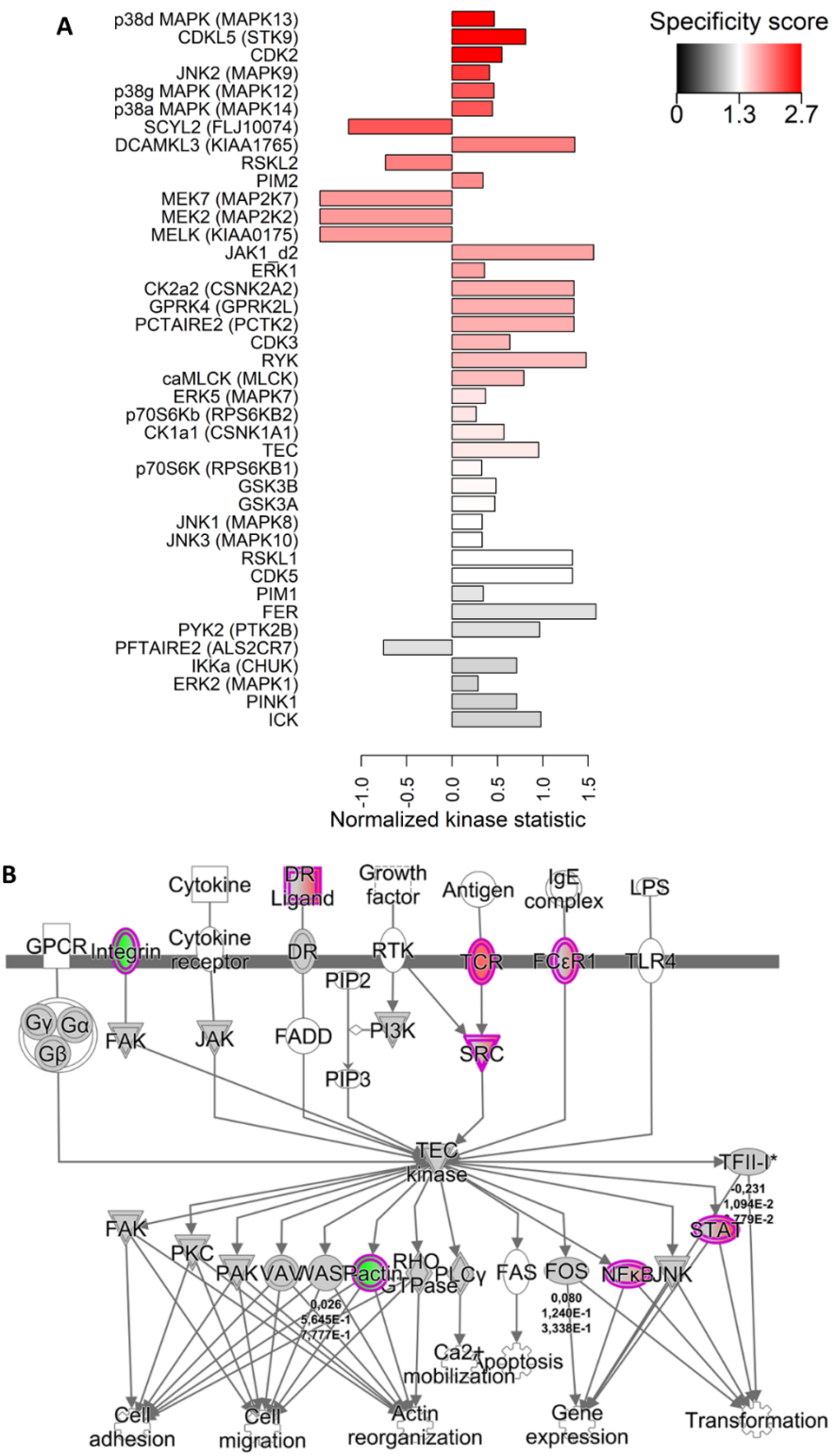
IFN signaling is usually stimulated by activated pattern recognition receptors. At the transcriptional level, upregulation of protein kinase receptor (PKR), a cytoplasmatic pattern-recognition receptor could be observed, which triggers various antiviral effects especially the production of type I IFNs. Interferon regulatory factor 7 (IRF7), a key protein responsible for induction of IFN expression was also found strongly upregulated, which together with IRF3 regulates expression of early type I IFN and other proteins involved in the innate antiviral immune response (*activation of IRF by cytosolic pattern recognition receptors*) (**Supplementary tables 1-2-3**). Signal transduction via PKR occurs mainly via NF-κB and MAPK pathways (‘*Role of PKR in Interferon induction and antiviral response’*). Another important intracellular pattern-recognition receptor for viral RNA which was found to be upregulated was MDA5. Furthermore, upregulation of the NF-κB subunits RelB and NFKB2/p52 can promote downstream production of in innate immunity chemokines (*NF-κB activation by viruses, NF-κB signaling*).

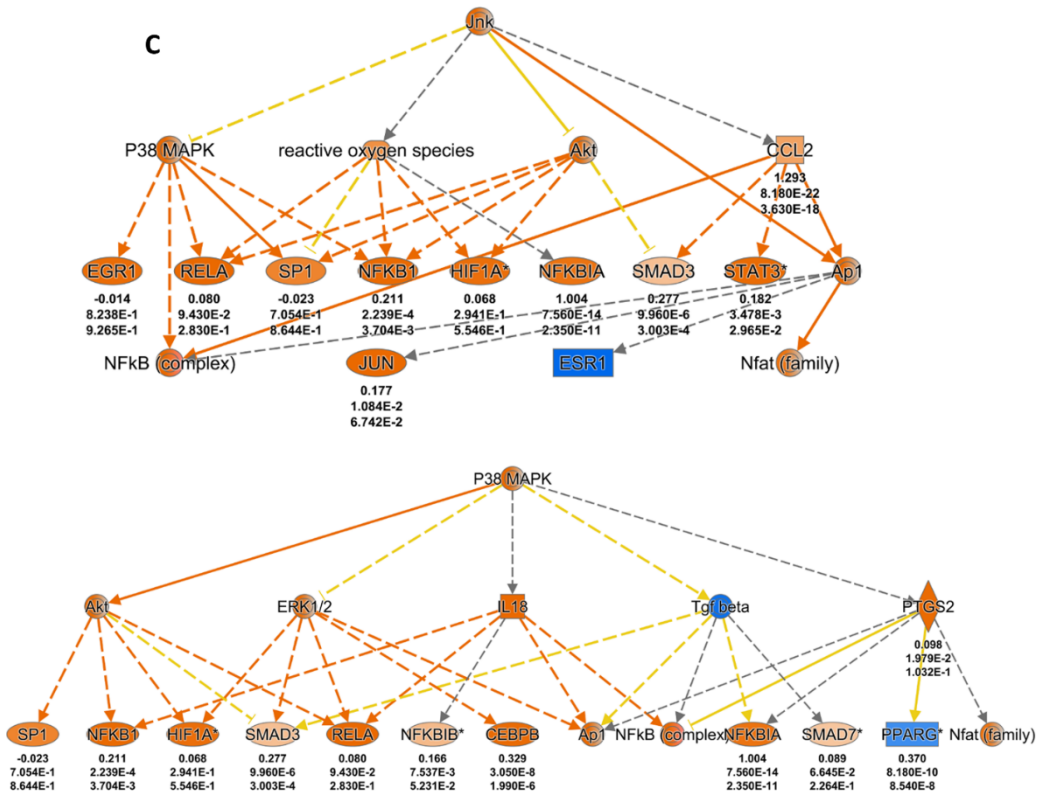
To validate upstream pathways which trigger downstream gene expression changes in THP1 monocytes following Echinaforce® treatment, we performed a Pamchip kinome activity profiling assay [38]. This peptide array approach allows characterization of cellular serine/threonine or tyrosine kinome activity profiles following on chip *in vitro* kinase reaction of 144 conserved kinase consensus peptide motifs in presence of THP1 monocyte lysates left untreated or following Echinaforce® treatment [56-59]. Using the upstream kinase prediction tool of the Bionavigator PamGene software, the qualitative and quantitative changes in phosphopeptide chip intensities upon Echinaforce® treatment were translated into a pattern of activated or inhibited upstream kinases (**Figure 3A and Supplementary table 4**). In agreement with the transcriptional activation of the IFN signaling pathway (**Figure 1C**), Pamchip kinome profiling reveals activation of the JAK1 kinase which is important in the phosphorylation of STAT kinases

and subsequently upregulation of IFN-stimulated genes. Furthermore, in line with pathway analysis of transcriptome data, we also identified activation of the tyrosine kinase TEC, known to phosphorylate and/or interact with JAK1 and JAK2 [60, 61] (**Figure 3B**) (**Supplementary table 2**). TEC was demonstrated to have important roles during innate immunity, i.e. TLR signaling [62], assembly and activation of the caspase-8 inflammasome [63], macrophage survival [64], IL-8 production [65], phagocytosis [66], and NF-kB signaling [67]. Besides, we also identified various Echinaforce® activated kinases belonging to the MAPK superfamily of kinases: p38 MAPK (MAPK11, -12, -13, and -14), JNK (MAPK8, -9 and -10) and ERK1, similar to be IPA-predicted upstream regulators based on the gene expression profiles (**Figure 3C**) related to various enriched canonical pathways: 'Role of Pattern Recognition Receptors in Recognition of Bacteria and Viruses', 'Activation of IRF by cytosolic Pattern Recognition Receptors', 'Dendritic cell Maturation', 'Role of MAPK Signaling in the Pathogenesis of Influenza', and 'IL-6 signaling', amongst others. In contrast, the activity of the MAP kinase kinases MAP2K7 and MAP2K2 were predicted to be suppressed. Various studies confirm involvement of these kinases in *Echinacea* biological action [20, 52, 68-71]. Alkylamides in the Echinaforce® tincture were found to be responsible for MAPK effects upon binding to CB2 receptors leading to increased cAMP, P38/MAPK and JNK signaling, NF-κB and ATF-2/CREB-1 activation [20].

Similarly, lipophilic extracts of *Echinacea* promoted murine dendritic cell maturation and mobility via the modulation of JNK, P38 MAPK and NF-κB pathways [52, 68, 69].

To further validate our kinase profiling, transcriptome and pathway analysis results showing crucial involvement of JAK kinase activation in IFN signaling upon Echinaforce® treatment, we compared THP1 gene expression changes following Echinaforce® treatment in presence or absence of the pharmacological JAK1 inhibitor Filgotinib. We found that Filgotinib significantly suppresses the Echinaforce® responsive genes MX1 and IFITM1, whereas STAT1, CXCL10 and IL8 gene expression were less significantly suppressed (**Figure 3D**). Altogether, our experiments with the JAK1 inhibitor Filgotinib strengthen our transcriptome and kinome data analysis, pointing to the critical involvement of JAK1-dependent IFN signaling in response to Echinaforce® treatment.


A
p38d MAPK (MAPK13)
CDKL5 (STK9)
CDK2
JNK2 (MAPK9)
p38g MAPK (MAPK12)
p38a MAPK (MAPK14)
SCYL2 (FLJ10074)
DCAMKL3 (KIAA1765)
RSKL2
PIM2
MEK7 (MAP2K7)
MEK2 (MAP2K2)
MELK (KIAA0175)
JAK1_d2
ERK1
CK2a2 (CSNK2A2)
GPRK4 (GPRK2L)
PCTAIRE2 (PCTK2)
CDK3
RYK
caMLCK (MLCK)
ERK5 (MAPK7)
p70S6Kb (RPS6KB2)
CK1a1 (CSNK1A1)
TEC
p70S6K (RPS6KB1)
GSK3B
GSK3A
JNK1 (MAPK8)
JNK3 (MAPK10)
RSKL1
CDK5
PIM1
FER
PYK2 (PTK2B)
PFTAIRE2 (ALS2CR7)
IKKa (CHUK)
ERK2 (MAPK1)
PINK1
ICK
Specificity score
0
1.3
2.7
-1.0
-0.5
0.0
0.5
1.0
1.5
Normalized kinase statistic
B
Cytokine
DR Ligand
Growth factor
Antigen
IgE complex
LPS
GPCR
Integrin
Cytokine receptor
DR
RTK
TCR
FCεR1
TLR4
PIP2
Gγ
Gα
Gβ
FAK
JAK
FADD
PI3K
SRC
PIP3
TEC kinase
TFII-I*
-0,231
1,094E-2
-,779E-2
STAT
FAK
PKC
PAK
VAV
WAS
Pactin
RHO GTPase
PLCγ
FAS
FOS
NFκB
JNK
0,026
-5,645E-1
-,777E-1
0,080
1,240E-1
3,338E-1
Ca2+ mobilization
Apoptosis
Cell adhesion
Cell migration
Actin reorganization
Gene expression
Transformation


C
Jnk
P38 MAPK
reactive oxygen species
Akt
CCL2
1.293
8.180E-22
3.630E-18
EGR1
-0.014
8.238E-1
9.265E-1
RELA
0.080
9.430E-2
2.830E-1
SP1
-0.023
7.054E-1
8.644E-1
NFKB1
0.211
2.239E-4
3.704E-3
HIF1A*
0.068
2.941E-1
5.546E-1
NFKBIA
1.004
7.560E-14
2.350E-11
SMAD3
0.277
9.960E-6
3.003E-4
STAT3*
0.182
3.478E-3
2.965E-2
Ap1
NFkB (complex)
JUN
0.177
1.084E-2
6.742E-2
ESR1
Nfat (family)
P38 MAPK
Akt
ERK1/2
IL18
Tgf beta
PTGS2
0.098
1.979E-2
1.032E-1
SP1
-0.023
7.054E-1
8.644E-1
NFKB1
0.211
2.239E-4
3.704E-3
HIF1A*
0.068
2.941E-1
5.546E-1
SMAD3
0.277
9.960E-6
3.003E-4
RELA
0.080
9.430E-2
2.830E-1
NFKBIB*
0.166
7.537E-3
5.231E-2
CEBPB
0.329
3.050E-8
1.990E-6
Ap1
NFkB (complex)
NFKBIA
1.004
7.560E-14
2.350E-11
SMAD7*
0.089
6.645E-2
2.264E-1
PPARG*
0.370
8.180E-10
8.540E-8
Nfat (family)

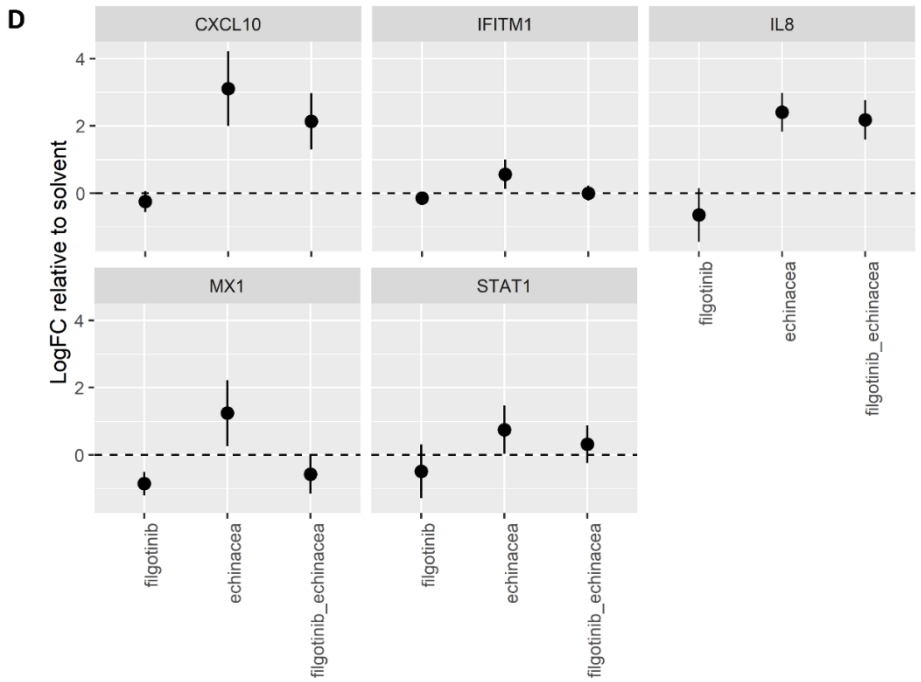

D
CXCL10
IFITM1
IL8
MX1
STAT1
LogFC relative to solvent
4
2
0
filgotinib
echinacea
filgotinib_echinacea

Figure 3: Echinaforce® activates JAK1 and MAPK kinases. A) PamGene kinome profiling of THP1 cell lysates, following 15 min treatment with Echinaforce® tincture (1%). Predicted upstream kinases. Bars are colored by specificity score. The direction of the bars represents the normalized kinase statistics. A positive kinase statistic means a higher activity in Echinaforce® treated samples. **B)** IPA TEC signaling pathway with Echinaforce®-induced up- and downregulated genes colored in red and green, respectively. TEC and JAK1 kinases were found to be activated by Echinaforce® treatment. **C)** IPA P38 MAPK and JNK upstream regulators. Genes colored in orange are predicted to be activated, while genes colored in blue are predicted to be inhibited. **D)** Effect of JAK1 inhibition on interferon qPCR gene expression response. THP1 cells were either treated with the JAK1 inhibitor Filgotinib alone or in combination with Echinaforce® (n=7). Mean expression LogFC change relative to solvent control is represented together with 95% confidence interval.

Echinaforce® treatment triggers DNA hypermethylation in CpG-poor gene bodies, in LINE, SINE and LTR transposon DNA repeat elements and silences endogenous retroviral sequences (ERVs)

Next, we characterized epigenetic changes following Echinacea treatment of THP1 cells. Different phytochemicals and nutrients are known to change DNA methylation and histone modifications by directly influencing epigenetic enzymes or by interfering with the availability of the substrates/cofactors of these enzymes [72-74]. In addition, epigenetics seems to be important during monocyte differentiation and in the immunological memory of macrophages [75, 76]. Today, various bioactive phytochemicals have been identified which modulate inflammation through epigenetic reprogramming [77, 78]. To assess whether the Echinaforce®-induced changes in transcriptome profiles in THP1 cells are associated with DNA methylation changes, we measured complementary changes in DNA methylation profiles using the Illumina EPIC methylation array. Significant DNA methylation changes were observed following 48h exposure to Echinaforce® **(Figure 4A and Supplementary table 5)**. 1,875 CpG sites were found differentially methylated (FDR < 0.1) with a methylation difference of at least 5%. Typically, DNA methylation changes following short exposures (24-72h) to phytochemicals and nutrients report are much smaller [45] than cancer associated DNA methylation changes in oncogenes or tumor suppressor genes which accumulate for many years in response to the microenvironment [79]. However, similar DMR effects sizes and cutoff (<5%) have been shown to be biologically meaningful in various disease etiologies in different DNA methylation studies [80, 81] [39].

From the 1,875 CpG sites identified, only 40 differentially methylated positions (DMPs) were hypomethylated whereas 1,835 DMPs were hypermethylated. DMPs were mainly enriched in gene bodies, intergenic, and CpG-poor regions, while depleted in CpG islands, promoter, and enhancer regions (**Figure 4D**). Only 1,259 of the 1,875 CpG-probes (67%) were located in a gene or 1,500 bp upstream of a gene. Similarly, DNA methylation variation in the immune system was predominantly found at CG islands

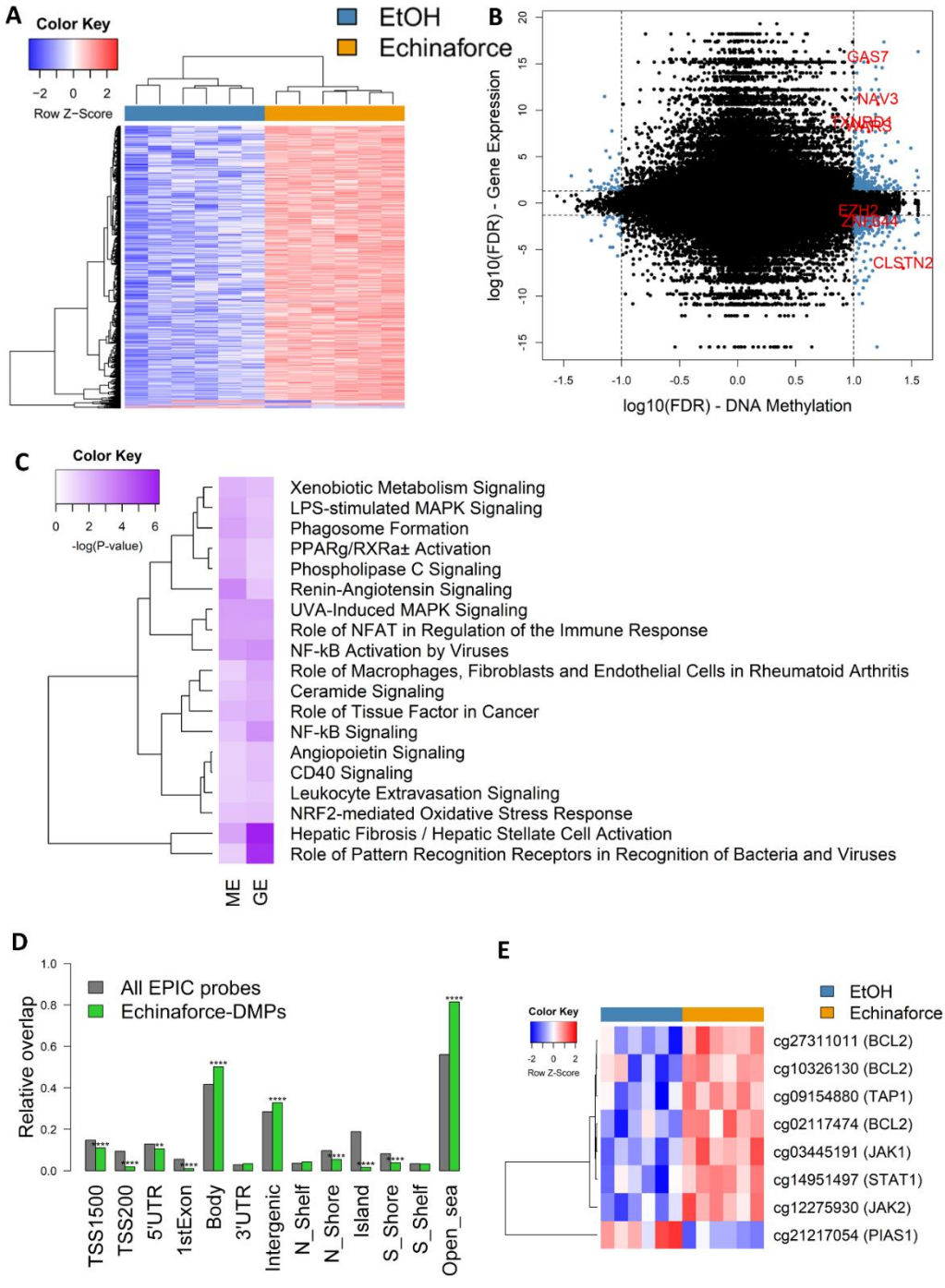
within gene bodies, which have the properties of cell type-restricted promoters, but infrequently at annotated gene promoters or CGI flanking sequences (CGI "shores") [82].

Since both gene expression and kinase profiling both revealed the involvement of interferon signaling pathways, we also checked whether methylation of IFN pathway genes was affected by Echinaforce® treatment. Eight probes located in BCL2, JAK1, STAT1, PIAS1 and TAP1 did show an FDR < 0.1, with weak methylation differences (between 1 and 3%). Whether this small methylation changes are sufficient to finetune the immune gene response needs further investigation by epigenetic editing approaches (**Figure 4E**). For example, the Echinacea induced IFN response may trigger a memory response by priming chromatin to mount faster and higher innate immune transcription upon re-stimulation of immune cells [83]. Along the same line, epigenetic bioactivities of Echinacea constituents may promote cryptic transcription of treatment-induced non-annotated TSSs (TINATs) encoding immunogenic peptides which might prime an antiviral innate immune response [84]. Furthermore, IPA pathway enrichment analysis of the genes containing DMPs revealed neuroinflammatory pathways as well as inflammation or immunological diseases (**Supplementary Table 6).** Of particular interest, one of the top enriched pathways ('Superpathway of Inositol Phosphate Compounds') controls various epigenetic processes and is critically involved in the interferon response [85-87].

Since most of the DMPs were located in intergenic regions and gene bodies, only a small subset of genes containing a DMP also resulted in a significant change in gene expression (**Figure 4B**). Only seven genes were both differentially methylated and expressed, based on the significance criteria described above: i.e. CLSTN2, EZH2, GAS7, NAV3, TXNRD1, WARS and ZNF644. When using less stringent significance criteria, leaving out the effect size cutoff (logFC), 574 CpG site – gene pairs were found to be differentially expressed and methylated. Upon further comparing canonical pathways which are significantly enriched for both lists of differentially expressed genes as well as the list of differentially methylated genes, we identified 10 common biological processes (**Figure 4C**). Remarkably, common pathways include NF-κB signaling (NF-κB activation by viruses, NF-κB signaling), MAPK signaling (LPS-stimulated MAPK signaling, UVA-induced MAPK signaling), and immune responses (Role of pattern recognition receptors in recognition of bacteria and viruses, Role of NFAT in regulation of the immune response, phagosome formation, CD40 signaling, leukocyte extravasation signaling). Finally, few Echinaforce®-induced DMPs overlap with immune cell (monocyte, macrophage) differentiation DNA methylation markers obtained from the BLUEPRINT consortium [51] (data not shown). Altogether, Echinaforce® treatment may

promote weak epigenetic changes on innate immune response but not adaptive immune response genes.

Beside the regulation of gene expression, DNA methylation is also important in alternative splicing, intron retention and maintaining genomic stability to prevent mobility of transposon repeats [88-92]. The higher global methylation seen in these repeat elements after Echinaforce® treatment may therefore contribute to genomic stability and preventing transposon activity. Interestingly, LINE-1 transposons can induce an interferon response which prevents further retrotransposition [93, 94]. Along the same line, it has been reported that the IFN response maintains DNA methylation silencing of repeats and noncoding RNAs [95]. Similarly, epigenetic DNA methylation drugs can trigger an IFN response through viral mimicry via transcription of dsRNAs derived at least in part by RNA polymerase III-driven bi-directional transcription of repetitive elements from endogenous retroviral elements [35, 93, 96]. In this respect, we also checked the methylation changes at repeat and transposon elements. These DNA repeats require hypermethylation to maintain genomic instability and prevent transposition. DMPs were found to be enriched in LINE, SINE and LTR transposon repeats, flanking endogenous retroviral sequences (ERVs) (**Figure 4F**). ERV sequences are major contributors in shaping and expanding the interferon network [97]. RNA transcripts of ERVs can be reverse transcribed to generate ssDNA or expressed to generate proteins with viral signatures, much like the pathogen-associated molecular patterns of exogenous viruses, which allows them to be detected by the innate immune system [98]. ERV expressed products have been shown to modulate innate immunity effectors, being therefore often related on the one side to inflammatory and autoimmune disorders, while on the other side to the control of excessive immune activation through their immunosuppressive properties. Finally, specific ERVs have been proposed to establish a protective effect against exogenous viral infections [97]. Interestingly, various DMR repeat sequences, flanking ERVs showed a global hypermethylation upon Echinaforce® treatment (**Figure 4G**). Along the same line, we observed decreased transcription of MER4D, MER57B1, MLT1C627, MLT2B4 ERVs after 12 and 48h Echinacea treatment, whereas MLT1B and MLT1C49 ERVs were transiently repressed at 12h but not at 48h. These results suggest that Echinaforce® triggers an evolutionary conserved (epi)genomic protective response against viral infection, upon hypermethylation of DNA repeats and silencing of flanking ERVs (**Figure 4H**).


A
Color Key
−2 0 2
Row Z−Score
EtOH
Echinaforce
B
log10(FDR) - Gene Expression
log10(FDR) - DNA Methylation
GAS7
NAV3
EZH2
CLSTN2
C
Color Key
0 2 4 6
-log(P-value)
Xenobiotic Metabolism Signaling
LPS-stimulated MAPK Signaling
Phagosome Formation
PPARg/RXRa± Activation
Phospholipase C Signaling
Renin-Angiotensin Signaling
UVA-Induced MAPK Signaling
Role of NFAT in Regulation of the Immune Response
NF-kB Activation by Viruses
Role of Macrophages, Fibroblasts and Endothelial Cells in Rheumatoid Arthritis
Ceramide Signaling
Role of Tissue Factor in Cancer
NF-kB Signaling
Angiopoietin Signaling
CD40 Signaling
Leukocyte Extravasation Signaling
NRF2-mediated Oxidative Stress Response
Hepatic Fibrosis / Hepatic Stellate Cell Activation
Role of Pattern Recognition Receptors in Recognition of Bacteria and Viruses
ME
GE
D
All EPIC probes
Echinaforce-DMPs
Relative overlap
TSS1500
TSS200
5'UTR
1stExon
Body
3'UTR
Intergenic
N_Shelf
N_Shore
Island
S_Shore
S_Shelf
Open_sea
E
Color Key
Row Z-Score
EtOH
Echinaforce
cg27311011 (BCL2)
cg10326130 (BCL2)
cg09154880 (TAP1)
cg02117474 (BCL2)
cg03445191 (JAK1)
cg14951497 (STAT1)
cg12275930 (JAK2)
cg21217054 (PIAS1)

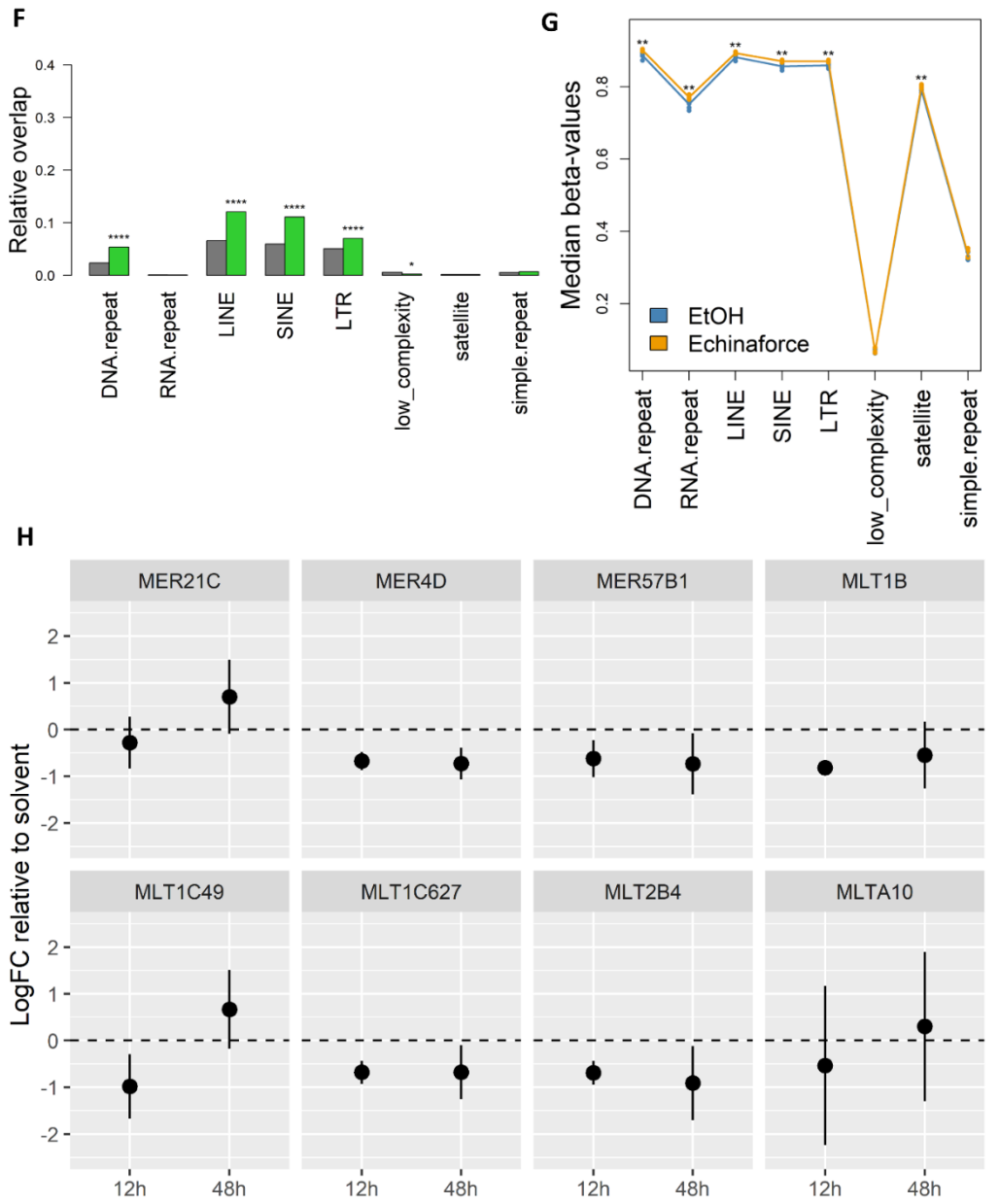


Figure 4: Echinaforce® treatment leads to global hypermethylation in CpG-poor gene bodies and intergenic repeat elements. A) Heatmap showing the methylation values of differentially methylated probes upon treatment of THP1 cells for 48h with Echinaforce® tincture (1%). Solvent (EtOH) controls are colored in blue and Echinaforce®-treated cells in orange. **B)** Starburst plot showing the genes both differentially expressed and differentially methylated. Each CpG-probe was mapped to its corresponding gene and the -log10(FDR) from the gene expression and DNA methylation analysis is displayed. The –log10(FDR) values of genes or CpG-probes with a negative LogFC or delta beta was multiplied by -1 leading to positive values when logFC or delta beta was positive and negative values when logFC or delta beta was negative. CpG-probe – gene pairs which were differentially expressed (FDR < 0.05) and differentially methylated (FDR < 0.1) were colored in blue. The CpG-probe – gene pairs of which the absolute delta beta was higher than 0.05 and the absolute logFC higher than 0.4 were colored in red. **C)** The IPA canonical

pathways which were both significantly enriched in the gene expression and DNA methylation analysis. **D)** Genomic enrichment of DMPs in different genomic regions. **E)** CpG probes located in genes of the interferon signaling pathway which were differentially methylated (FDR < 0.1). * P ≤ 0.05, ** P ≤ 0.01, *** P ≤ 0.001, **** P ≤ 0.0001. **F)** Genomic enrichment of DMPs in different repeat elements. GE) Global DNA methylation in different repeat elements. **H)** hERV qPCR gene expression. THP1 cells were with Echinaforce® at 12 and 48h (n=3). Mean LogFC change relative to solvent control is represented together with 95% confidence interval.

Conclusion

In this study, we applied a systems biology approach to characterize immunological pathways targeted by a widely used standardized ethanolic *Echinacea purpurea* tincture (Echinaforce®, A.Vogel Bioforce, Switzerland). Microarray and QPCR based assays revealed that Echinaforce® treatment induces time dependent changes in innate immunity related gene expression, involving IFN and chemokine innate immune signaling in THP1 monocytic cells. Based on phosphopeptide based kinome data analysis and pharmacological inhibitor experiments with the specific JAK1 inhibitor Filgotinib, we further demonstrate crucial involvement of JAK1 activation in IFN and immunity related gene activation by Echinaforce treatment. Finally, DNA hypermethylation changes following Echinaforce® treatment were predominantly observed in intergenic CGIs, genomic DNA repeat elements (LINE, SINE, LTR) flanking ERVs and to a less extent in regulatory promoter regions, favoring activation of an evolutionary conserved (epi)genomic protective response against viral infection. The functional role of these methylation changes in immunogenic priming of innate immunity, ERV silencing and genome integrity needs to be further investigated.

Altogether, we show that Echinaforce® treatment of THP1 cells activates an innate immune response by priming JAK1 dependent interferon signaling and chemokine gene expression, DNA repeat hypermethylation and silencing of flanking endogenous retroviral sequences, which might contribute in protection against respiratory infections. Finally, immunoprotective effects of Echinaforce® need to be further validated in primary immune cell types, in preclinical respiratory infection models or placebo controlled intervention studies.

References

1. Barnes, J., et al., *Echinacea species (Echinacea angustifolia (DC.) Hell., Echinacea pallida (Nutt.) Nutt.,Echinacea purpurea (L.) Moench): a review of their chemistry, pharmacology and clinical properties.* J Pharm Pharmacol, 2005. **57**(8): p. 929-54.
2. Barrett, B., *Medicinal properties of Echinacea: a critical review.* Phytomedicine, 2003. **10**(1): p. 66-86.
3. Manayi, A., M. Vazirian, and S. Saeidnia, *Echinacea purpurea: Pharmacology, phytochemistry and analysis methods.* Pharmacogn Rev, 2015. **9**(17): p. 63-72.
4. Karsch-Volk, M., B. Barrett, and K. Linde, *Echinacea for preventing and treating the common cold.* JAMA, 2015. **313**(6): p. 618-9.
5. Binns, S.E., et al., *Phytochemical variation in echinacea from roots and flowerheads of wild and cultivated populations.* J Agric Food Chem, 2002. **50**(13): p. 3673-87.
6. Schapowal, A., P. Klein, and S.L. Johnston, *Echinacea reduces the risk of recurrent respiratory tract infections and complications: a meta-analysis of randomized controlled trials.* Adv Ther, 2015. **32**(3): p. 187-200.
7. Hudson, J.B., *Applications of the phytomedicine Echinacea purpurea (Purple Coneflower) in infectious diseases.* J Biomed Biotechnol, 2012. **2012**: p. 769896.
8. Chicca, A., et al., *Synergistic immunomopharmacological effects of N-alkylamides in Echinacea purpurea herbal extracts.* Int Immunopharmacol, 2009. **9**(7-8): p. 850-8.
9. Dalby-Brown, L., et al., *Synergistic antioxidative effects of alkamides, caffeic acid derivatives, and polysaccharide fractions from Echinacea purpurea on in vitro oxidation of human low-density lipoproteins.* J Agric Food Chem, 2005. **53**(24): p. 9413-23.
10. Benson, J.M., et al., *Echinacea purpurea extracts modulate murine dendritic cell fate and function.* Food Chem Toxicol, 2010. **48**(5): p. 1170-7.
11. Ritchie, M.R., et al., *Effects of Echinaforce(R) treatment on ex vivo-stimulated blood cells.* Phytomedicine, 2011. **18**(10): p. 826-31.
12. Sharma, M., et al., *Induction of multiple pro-inflammatory cytokines by respiratory viruses and reversal by standardized Echinacea, a potent antiviral herbal extract.* Antiviral Res, 2009. **83**(2): p. 165-70.
13. Sharma, S.M., et al., *Bactericidal and anti-inflammatory properties of a standardized Echinacea extract (Echinaforce): dual actions against respiratory bacteria.* Phytomedicine, 2010. **17**(8-9): p. 563-8.
14. Sharma, M., R. Schoop, and J.B. Hudson, *Echinacea as an antiinflammatory agent: the influence of physiologically relevant parameters.* Phytother Res, 2009. **23**(6): p. 863-7.
15. Pleschka, S., et al., *Anti-viral properties and mode of action of standardized Echinacea purpurea extract against highly pathogenic avian influenza virus (H5N1, H7N7) and swine-origin H1N1 (S-OIV).* Virol J, 2009. **6**: p. 197.
16. Sharma, M., R. Schoop, and J.B. Hudson, *The efficacy of Echinacea in a 3-D tissue model of human airway epithelium.* Phytother Res, 2010. **24**(6): p. 900-4.
17. Canlas, J., et al., *Echinacea and trypanasomatid parasite interactions: growth-inhibitory and anti-inflammatory effects of Echinacea.* Pharm Biol, 2010. **48**(9): p. 1047-52.
18. Woelkart, K., et al., *Bioavailability and pharmacokinetics of Echinacea purpurea preparations and their interaction with the immune system.* Int J Clin Pharmacol Ther, 2006. **44**(9): p. 401-8.
19. Sharma, M., et al., *The potential use of Echinacea in acne: control of Propionibacterium acnes growth and inflammation.* Phytother Res, 2011. **25**(4): p. 517-21.
20. Gertsch, J., et al., *Echinacea alkylamides modulate TNF-alpha gene expression via cannabinoid receptor CB2 and multiple signal transduction pathways.* FEBS Lett, 2004. **577**(3): p. 563-9.
21. Vimalanathan, S., et al., *Prevention of influenza virus induced bacterial superinfection by standardized Echinacea purpurea, via regulation of surface receptor expression in human bronchial epithelial cells.* Virus Res, 2017. **233**: p. 51-59.

22. Vimalanathan, S., R. Schoop, and J. Hudson, *High-potency Anti-influenza Therapy by a Combination of Echinacea purpurea fresh herb and root tinctures.* Journal of Applied Pharmaceutical Science 2013. **3**: p. 1-5.
23. Brinkeborn, R.M., D.V. Shah, and F.H. Degenring, *Echinaforce and other Echinacea fresh plant preparations in the treatment of the common cold. A randomized, placebo controlled, double-blind clinical trial.* Phytomedicine, 1999. **6**(1): p. 1-6.
24. Jawad, M., et al., *Safety and Efficacy Profile of Echinacea purpurea to Prevent Common Cold Episodes: A Randomized, Double-Blind, Placebo-Controlled Trial.* Evid Based Complement Alternat Med, 2012. **2012**: p. 841315.
25. Raduner, S., et al., *Alkylamides from Echinacea are a new class of cannabinomimetics. Cannabinoid type 2 receptor-dependent and -independent immunomodulatory effects.* J Biol Chem, 2006. **281**(20): p. 14192-206.
26. Bruni, R., et al., *Analytical methods for the study of bioactive compounds from medicinally used Echinacea species.* J Pharm Biomed Anal, 2018. **160**: p. 443-477.
27. Pugh, N.D., C.R. Jackson, and D.S. Pasco, *Total bacterial load within Echinacea purpurea, determined using a new PCR-based quantification method, is correlated with LPS levels and in vitro macrophage activity.* Planta Med, 2013. **79**(1): p. 9-14.
28. Pugh, N.D., et al., *The majority of in vitro macrophage activation exhibited by extracts of some immune enhancing botanicals is due to bacterial lipoproteins and lipopolysaccharides.* Int Immunopharmacol, 2008. **8**(7): p. 1023-32.
29. Tamta, H., et al., *Variability in in vitro macrophage activation by commercially diverse bulk echinacea plant material is predominantly due to bacterial lipoproteins and lipopolysaccharides.* J Agric Food Chem, 2008. **56**(22): p. 10552-6.
30. Todd, D.A., et al., *Ethanolic Echinacea purpurea Extracts Contain a Mixture of Cytokine-Suppressive and Cytokine-Inducing Compounds, Including Some That Originate from Endophytic Bacteria.* PLoS One, 2015. **10**(5): p. e0124276.
31. Dunning, M.J., et al., *beadarray: R classes and methods for Illumina bead-based data.* Bioinformatics, 2007. **23**(16): p. 2183-4.
32. Barbosa-Morais, N.L., et al., *A re-annotation pipeline for Illumina BeadArrays: improving the interpretation of gene expression data*. Nucleic Acids Res, 2010. **38**(3): p. e17.
33. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and microarray studies.* Nucleic Acids Res, 2015. **43**(7): p. e47.
34. Dunning M, L.A.a.E.M., *illuminaHumanv4.db: Illumina HumanHT12v4 annotation data (chip illuminaHumanv4)*. 2015, R package version 1.26.0.
35. Roulois, D., et al., *DNA-Demethylating Agents Target Colorectal Cancer Cells by Inducing Viral Mimicry by Endogenous Transcripts.* Cell, 2015. **162**(5): p. 961-73.
36. Reiner, A., D. Yekutieli, and Y. Benjamini, *Identifying differentially expressed genes using false discovery rate controlling procedures.* Bioinformatics, 2003. **19**(3): p. 368-75.
37. Safaei, J., et al., *Prediction of 492 human protein kinase substrate specificities.* Proteome Sci, 2011. **9 Suppl 1**: p. S6.
38. Chirumamilla, C.S., et al., *Profiling Activity of Cellular Kinases in Migrating T-Cells.* Methods Mol Biol, 2019. **1930**: p. 99-113.
39. Szarc Vel Szic, K., et al., *Epigenetic silencing of triple negative breast cancer hallmarks by Withaferin A.* Oncotarget, 2017. **8**(25): p. 40434-40453.
40. Assenov, Y., et al., *Comprehensive analysis of DNA methylation data with RnBeads.* Nat Methods, 2014. **11**(11): p. 1138-1140.
41. Teschendorff, A.E., et al., *A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data.* Bioinformatics, 2013. **29**(2): p. 189-96.
42. Hansen KD, *IlluminaHumanMethylationEPICmanifest: Manifest for Illumina's EPIC methylation arrays*. 2016, R package version 0.3.0.
43. Milenkovic, D., et al., *Dietary flavanols modulate the transcription of genes associated with cardiovascular pathology without changes in their DNA methylation state.* PLoS One, 2014. **9**(4): p. e95527.

44. Chirumamilla, C.S., et al., *Selective Glucocorticoid Receptor Properties of GSK866 Analogs with Cysteine Reactive Warheads.* Front Immunol, 2017. **8**: p. 1324.
45. Milenkovic, D., et al., *A systems biology network analysis of nutri(epi)genomic changes in endothelial cells exposed to epicatechin metabolites.* Sci Rep, 2018. **8**(1): p. 15487.
46. Verhelst, J., et al., *Interferon-inducible protein Mx1 inhibits influenza virus by interfering with functional viral ribonucleoprotein complex assembly.* J Virol, 2012. **86**(24): p. 13445-55.
47. Verhelst, J., P. Hulpiau, and X. Saelens, *Mx proteins: antiviral gatekeepers that restrain the uninvited.* Microbiol Mol Biol Rev, 2013. **77**(4): p. 551-66.
48. Smith, S.E., et al., *Interferon-Induced Transmembrane Protein 1 Restricts Replication of Viruses That Enter Cells via the Plasma Membrane.* J Virol, 2019. **93**(6).
49. Sokol, C.L. and A.D. Luster, *The chemokine system in innate immunity.* Cold Spring Harb Perspect Biol, 2015. **7**(5).
50. Chistiakov, D.A., et al., *The impact of interferon-regulatory factors to macrophage differentiation and polarization into M1 and M2.* Immunobiology, 2018. **223**(1): p. 101-111.
51. Adams, D., et al., *BLUEPRINT to decode the epigenetic signature written in blood.* Nat Biotechnol, 2012. **30**(3): p. 224-6.
52. Yin, S.Y., et al., *Stimulatory effect of Echinacea purpurea extract on the trafficking activity of mouse dendritic cells: revealed by genomic and proteomic analyses.* BMC Genomics, 2010. **11**: p. 612.
53. Sharma, M., J.T. Arnason, and J.B. Hudson, *Echinacea extracts modulate the production of multiple transcription factors in uninfected cells and rhinovirus-infected cells.* Phytother Res, 2006. **20**(12): p. 1074-9.
54. Cech, N.B., et al., *Echinacea and its alkylamides: effects on the influenza A-induced secretion of cytokines, chemokines, and PGE(2) from RAW 264.7 macrophage-like cells.* Int Immunopharmacol, 2010. **10**(10): p. 1268-78.
55. Sharma, M., et al., *Echinacea extracts modulate the pattern of chemokine and cytokine secretion in rhinovirus-infected and uninfected epithelial cells.* Phytother Res, 2006. **20**(2): p. 147-52.
56. Arsenault, R., P. Griebel, and S. Napper, *Peptide arrays for kinome analysis: new opportunities and remaining challenges.* Proteomics, 2011. **11**(24): p. 4595-609.
57. Dussaq, A., et al., *Mechanistic Parameterization of the Kinomic Signal in Peptide Arrays.* J Proteomics Bioinform, 2016. **9**(5): p. 151-157.
58. Labots, M., et al., *Evaluation of a tyrosine kinase peptide microarray for tyrosine kinase inhibitor therapy selection in cancer.* Exp Mol Med, 2016. **48**(12): p. e279.
59. Baharani, A., et al., *Technological advances for interrogating the human kinome.* Biochem Soc Trans, 2017. **45**(1): p. 65-77.
60. Takahashi-Tezuka, M., et al., *Tec tyrosine kinase links the cytokine receptors to PI-3 kinase probably through JAK.* Oncogene, 1997. **14**(19): p. 2273-82.
61. Yamashita, Y., et al., *Tec and Jak2 kinases cooperate to mediate cytokine-driven activation of c-fos transcription.* Blood, 1998. **91**(5): p. 1496-507.
62. Tampella, G., et al., *The Tec Kinase-Regulated Phosphoproteome Reveals a Mechanism for the Regulation of Inhibitory Signals in Murine Macrophages.* J Immunol, 2015. **195**(1): p. 246-56.
63. Zwolanek, F., et al., *The non-receptor tyrosine kinase Tec controls assembly and activity of the noncanonical caspase-8 inflammasome.* PLoS Pathog, 2014. **10**(12): p. e1004525.
64. Melcher, M., et al., *Essential roles for the Tec family kinases Tec and Btk in M-CSF receptor signaling pathways that regulate macrophage survival.* J Immunol, 2008. **180**(12): p. 8048-56.
65. Wang, G.Q., et al., *Tec kinase mediating IL-8 transcription in monocytes stimulated with LPS.* Inflammation, 2009. **32**(4): p. 265-9.
66. Jongstra-Bilen, J., et al., *Dual functions of Bruton's tyrosine kinase and Tec kinase during Fcgamma receptor-induced signaling and phagocytosis.* J Immunol, 2008. **181**(1): p. 288-98.
67. Wang, F., et al., *Inhibitor of Tec kinase, LFM-A13, decreases pro-inflammatory mediators production in LPS-stimulated RAW264.7 macrophages via NF-kappaB pathway.* Oncotarget, 2017. **8**(21): p. 34099-34110.
68. Wang, C.Y., et al., *Genomics and proteomics of immune modulatory effects of a butanol fraction of echinacea purpurea in human dendritic cells.* BMC Genomics, 2008. **9**: p. 479.

69. Li, Y., et al., *Echinacea pupurea extracts promote murine dendritic cell maturation by activation of JNK, p38 MAPK and NF-kappaB pathways.* Dev Comp Immunol, 2017. **73**: p. 21-26.
70. Sullivan, A.M., et al., *Echinacea-induced macrophage activation.* Immunopharmacol Immunotoxicol, 2008. **30**(3): p. 553-74.
71. Fu, A., et al., *Echinacea purpurea Extract Polarizes M1 Macrophages in Murine Bone Marrow-Derived Macrophages Through the Activation of JNK.* J Cell Biochem, 2017. **118**(9): p. 2664-2671.
72. Vanden Berghe, W., *Epigenetic impact of dietary polyphenols in cancer chemoprevention: lifelong remodeling of our epigenomes.* Pharmacol Res, 2012. **65**(6): p. 565-76.
73. Milagro, F.I., et al., *Dietary factors, epigenetic modifications and obesity outcomes: progresses and perspectives.* Mol Aspects Med, 2013. **34**(4): p. 782-812.
74. Remely, M., et al., *Therapeutic perspectives of epigenetically active nutrients.* Br J Pharmacol, 2015. **172**(11): p. 2756-68.
75. Novakovic, B., et al., *beta-Glucan Reverses the Epigenetic State of LPS-Induced Immunological Tolerance.* Cell, 2016. **167**(5): p. 1354-1368 e14.
76. Saeed, S., et al., *Epigenetic programming of monocyte-to-macrophage differentiation and trained innate immunity.* Science, 2014. **345**(6204): p. 1251086.
77. Szarc vel Szic, K., et al., *From inflammaging to healthy aging by dietary lifestyle choices: is epigenetics the key to personalized nutrition?* Clin Epigenetics, 2015. **7**: p. 33.
78. Vanden Berghe, W., et al., *Keeping up NF-kappaB appearances: epigenetic control of immunity or inflammation-triggered epigenetics.* Biochem Pharmacol, 2006. **72**(9): p. 1114-31.
79. Koch, A., et al., *Analysis of DNA methylation in cancer: location revisited.* Nat Rev Clin Oncol, 2018. **15**(7): p. 459-466.
80. Dayeh, T., et al., *Genome-wide DNA methylation analysis of human pancreatic islets from type 2 diabetic and non-diabetic donors identifies candidate genes that influence insulin secretion.* PLoS Genet, 2014. **10**(3): p. e1004160.
81. Langie, S.A.S., et al., *GLI2 promoter hypermethylation in saliva of children with a respiratory allergy.* Clin Epigenetics, 2018. **10**: p. 50.
82. Deaton, A.M., et al., *Cell type-specific DNA methylation at intragenic CpG islands in the immune system.* Genome Res, 2011. **21**(7): p. 1074-86.
83. Kamada, R., et al., *Interferon stimulation creates chromatin marks and establishes transcriptional memory.* Proc Natl Acad Sci U S A, 2018. **115**(39): p. E9162-E9171.
84. Brocks, D., et al., *DNMT and HDAC inhibitors induce cryptic transcription start sites encoded in long terminal repeats.* Nat Genet, 2017. **49**(7): p. 1052-1060.
85. Pulloor, N.K., et al., *Human genome-wide RNAi screen identifies an essential role for inositol pyrophosphates in Type-I interferon response.* PLoS Pathog, 2014. **10**(2): p. e1003981.
86. Kutateladze, T.G., *Histone deacetylation: IP4 is an epigenetic coregulator*. Nat Chem Biol, 2012. **8**(3): p. 230-1.
87. Burton, A., et al., *Inositol pyrophosphates regulate JMJD2C-dependent histone demethylation.* Proc Natl Acad Sci U S A, 2013. **110**(47): p. 18970-5.
88. Meng, H., et al., *DNA methylation, its mediators and genome integrity.* Int J Biol Sci, 2015. **11**(5): p. 604-17.
89. Papin, C., et al., *Combinatorial DNA methylation codes at repetitive elements.* Genome Res, 2017. **27**(6): p. 934-946.
90. Nanan, K.K., et al., *Independence between pre-mRNA splicing and DNA methylation in an isogenic minigene resource.* Nucleic Acids Res, 2017. **45**(22): p. 12780-12797.
91. Jones, P.A., *Functions of DNA methylation: islands, start sites, gene bodies and beyond.* Nat Rev Genet, 2012. **13**(7): p. 484-92.
92. Kim, D., et al., *Population-dependent Intron Retention and DNA Methylation in Breast Cancer.* Mol Cancer Res, 2018. **16**(3): p. 461-469.
93. Chiappinelli, K.B., et al., *Inhibiting DNA Methylation Causes an Interferon Response in Cancer via dsRNA Including Endogenous Retroviruses.* Cell, 2015. **162**(5): p. 974-86.
94. Yu, Q., et al., *Type I interferon controls propagation of long interspersed element-1.* J Biol Chem, 2015. **290**(16): p. 10191-9.

95. Leonova, K.I., et al., *p53 cooperates with DNA methylation and a suicidal interferon response to maintain epigenetic silencing of repeats and noncoding RNAs.* Proc Natl Acad Sci U S A, 2013. **110**(1): p. E89-98.
96. Licht, J.D., *DNA Methylation Inhibitors in Cancer Therapy: The Immunity Dimension.* Cell, 2015. **162**(5): p. 938-9.
97. Grandi, N. and E. Tramontano, *Human Endogenous Retroviruses Are Ancient Acquired Elements Still Shaping Innate Immune Responses.* Front Immunol, 2018. **9**: p. 2039.
98. Hurst, T.P. and G. Magiorkinis, *Activation of the innate immune response by endogenous retroviruses.* J Gen Virol, 2015. **96**(Pt 6): p. 1207-18.

Supplementary Information

Supplementary tables 1 to 6 can be found in following dropbox folder: https://www.dropbox.com/sh/d9lq9mipqkoezx3/AAC9RCXgWS7P6DS71ybJqWNJa?dl=0

Supplementary Table 1: Differentially expressed probes (FDR < 0.05 and logFC > 0.4) after Echinaforce® tincture treatment.

Supplementary Table 2: Enriched Ingenuity canonical pathways of differentially expressed genes after Echinaforce® tincture treatment.

Supplementary Table 3: Enriched Ingenuity diseases and bio functions of differentially expressed genes after Echinaforce® tincture treatment.

Supplementary Table 4: PamGene upstream kinase analysis.

Supplementary Table 5: DMPs (FDR < 0.05 and |DeltaBetas| > 0.05).

Supplementary Table 6: Enriched Ingenuity canonical pathways of differentially methylated genes after Echinaforce® tincture treatment.

Conflict of interest statement

Dr. A. Suter was employed by the company Dr. Vogel (Switzerland). The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

9

General conclusion & future perspectives

The main aim of this thesis was to evaluate the applicability of DNA methylation signatures from easy accessible tissues like blood and saliva to monitor past adverse environmental exposures, cardio-metabolic and inflammaging diseases and effects of dietary or phytomedicinal compounds. To what extent DNA methylation profiles allow us to monitor our health status and advise us for personalized lifestyle interventions? Which hurdles should still be overcome to fully exploit the potential of epigenetics in personalized health applications?

Can blood and saliva being used in epigenetic-based health prediction?

Tissue-specificity of DNA methylation changes

Blood and saliva are typically not the primary cell or tissue of interest when studying chronic lifestyle diseases (cancer, metabolic disorder, neurodegenerative disease, CVD). For example, in atherosclerosis, or Alzheimer's and Parkinson's disease, the diseased tissue of interest (e.g. plaques in the blood vessels, or brain) are usually not accessible and therefore blood and saliva may represent interesting surrogate tissues to monitor pathological systemic changes or to identify disease associated biomarkers. Also in metabolic disorders such as obesity and diabetes, tissues such as adipose, liver, pancreas and muscle may be more informative than blood and saliva. However, their ease of collection makes blood and saliva more useful for clinical and health diagnostic applications, and is therefore often the primary choice in epigenetic studies. This of course raises questions whether the DNA methylation changes observed in blood or saliva are relevant for the primary disease tissue of interest.

Although a lot of EWAS use peripheral blood to study disease mechanisms or detect biomarkers, information about sensitivity and specificity of these surrogate tissue biomarkers is scarce. Studies examining the comparability between matched blood and other tissue samples in general show a shift towards more positive correlations between the methylation values across the tissues. However, in general only a fraction of these sites is highly correlated ($r>0.5$) and considered as concordant CpG sites. In a study of 143 individuals with paired blood and adipose tissue methylation data, FAS gene methylation associations with BMI were only detected in adipose tissue [1]. In contrast, HIF3A gene methylation could be correlated with BMI in both blood and adipose tissue. Although this epigenetic association of HIF3A gene methylation across tissues was also confirmed in other studies [2], it could not be confirmed in skin tissue

[2] illustrating that this association is not completely tissue-independent. This suggests that blood-based markers may have diagnostic value. Also other sites were found to be differentially methylated in both blood and adipose tissue associated with lipid and cholesterol levels [3, 4], and obesity [5]. Similar results were reported when comparing blood with different brain tissues, where only a small fraction of the CpG sites were highly concordant, whereas other sites were not or rather weakly correlated. This limited concordance may therefore hamper the use of surrogate tissues to identify important molecular mechanistic insights. Some resources exist which can be used to check the concordance of individual sites across blood and brain tissue, and may be used to prioritize significant hits in blood studies for further functional validation [6, 7]. However, the low sample size of some studies may underestimate the real concordance. Another strategy is to only use concordant CpG sites in blood-based EWAS, which was recently successfully being performed in a study of schizophrenia [8]. Furthermore, studies in AD [9], PD [10] and schizophrenia [11] identified blood specific DMPs associated with DMP changes in brain tissues, opening new perspectives for blood methylome biomarkers in brain disorders.

Although saliva sampling is easier to implement than blood sampling in vulnerable populations in non-clinical settings [12, 13], only a limited set of studies performed methylation studies in saliva samples. Saliva and blood methylomes were found to show significant overlap [14-17]. Interestingly, Smith and colleagues observed that saliva methylation profiles were more similar to brain tissues than blood methylomes [17]. They speculated that this result could be the similar ectodermal origin of buccal cells and cells of the central nervous system. A similar conclusion could be drawn from a study which compared cord tissue with cord blood in a neonatal EWAS. Here, cord tissue was found to be a better surrogate for target tissues of mesodermal origin [18]. In the same view, inflammation/immune related diseases may be best captured using blood. Therefore, the choice of surrogate tissue type may also be dependent on the disease tissue of interest [19]. An interesting recent study obtained matched blood, saliva, buccal and live brain tissues, and concluded that the correlation between brain-peripheral tissue is also CpG and gene-dependent [20].

In chapter 5, we identified a hypermethylation in the BRCA1-NBR2 and CRISP2 promoter region, which was evident in both blood and atherosclerotic plaques, and which was not detected in other diseases, indicating the specificity of this marker. Despite replication, further replication is needed in bigger cohort studies before these markers can be applied as surrogate biomarkers. Although a common change in the surrogate tissue and tissue of interest may indicate the potential importance of this marker in the disease, it is not an absolute requirement for biomarker research. DNA methylation changes in blood or saliva which are not causal or which are not observed

in the target tissue, may still hold promise as prognostic or therapeutic biomarker, but are less applicable as a therapeutic target.

The problem of cellular heterogeneity

Another issue when studying genome-wide DNA methylation in blood and saliva, is the cellular heterogeneity of these tissues. Blood and saliva consist of different cell types with their own cell type-specific methylome. Therefore, a shift in relative cell type distribution due to disease or exposures may lead to an observable change in DNA methylation. When studying mechanistic epigenetic processes or identifying therapeutic targets in blood samples or other heterogenous tissues, it is crucial to take these cellular dynamics into account. A large proportion of DNA methylation variability in blood samples can be attributed to this cellular effect. For example, one of the first studies demonstrating the strong impact of cellular heterogeneity on EWAS outcome, was a genome-wide DNA methylation study in rheumatoid arthritis measured in whole blood. No correction for differential blood cell counts resulted in a high number of significant associations, while most of these associations disappeared upon cell type correction [21]. Another study demonstrated that most of the previously reported changes in DNA methylation during aging were due to shifts in cell type composition [22], illustrating the need of adjusting for cellular heterogeneity. Based on these early studies, multiple statistical methods have been designed to adjust for cell type effects in EWAS. With respect to EWAS in blood samples, the most popular method is the one proposed by Houseman and colleagues who made use of reference methylomes from six major blood cell types to estimate relative cell type contributions in blood samples. In this method, one assumes that the sample's DNA methylation profile is the result of a linear combination of the underlying cell type methylomes [23]. These estimated cell counts can then be used to correct the associations between methylation and the phenotype of interest.

This method is not restricted to blood, but can also be used for other tissues with known underlying cell type composition and available reference methylomes. In this way, several reference methylomes have been constructed for whole blood, cord blood, prefrontal cortex and breast tissue [24]. In chapter 6, we used reference methylomes from ENCODE to estimate the main cell type composition in plaques and aorta, and found that a large part of the methylation differences observed in atherosclerotic plaques can be attributed to immune cell infiltration. This indicates that one must be careful in interpreting EWAS studies of complex solid tissues, like aorta and plaques, where considerable shifts in cell type distribution can be expected.

Although I used the major cell types involved in atherosclerotic plaques to construct this reference methylome, plaques are much more complex and it is difficult to have a complete picture of all cell types and cellular activation states involved. For most tissues, the exact underlying cell type composition is not known and/or reference methylomes are not available. Therefore, also multiple reference-free algorithms have been constructed, such as EWASher, RefFreeEWAS and ReFACTOR, that may help in adjusting cellular effects without the need for reference methylomes [24]. Also in blood, different subtypes of CD4+ and CD8+ T-cells and activation states exist, for which reference methylomes are difficult to construct. For example, it was recently demonstrated that a highly replicated smoking-associated DNA hypomethylation in a CpG site in the GPR15 gene was not due to the direct effect of tobacco smoking compounds on DNA methylation, but was rather the consequence of an expansion of CD3+GPR15+ expression T cells [25]. This indicates that subtle cellular effects exist, and that there is a need to quantify blood cell type composition in more detail to elucidate the real extent of cellular heterogeneity. Consortia, like BLUEPRINT and the International Human Epigenome Consortium (IHEC) are now generating methylomes and epigenomes of multiple major and minor cell types, and may therefore help detecting and accounting for these rare cell subtypes in future blood EWAS. In addition, it may also contribute to the generation of new reference methylomes for complex solid tissues, such as brain, skin, lung etc., of which the underlying cellular components are less well known. Moreover, single-cell based epigenomic methods are on the rise to detect subtle cell type specific epigenetic effects, as recently demonstrated for CD4+ T-cell subtypes [26].

In theory, a detectable DNA methylation change measured in a complex tissue like blood and which is not the result of shifts in cell type distribution, can be due to a methylation change in a specific cell type, a subset of cell types or even all the underlying cell types. Knowledge about the cell type responsible for the methylation change may be useful in elucidating mechanistic pathways and interpretation of the results. Cell type separation techniques like FACS may, however, not distinguish rare cell subtypes [27], and furthermore, even purified cell types exhibit extensive variation in DNA methylation. Purified cell types were found to be rather a collection of different epigenomes or meta-epigenomes [28] and cell type effects may not completely be removed after cell type separation. Of interest, recently also a statistical tool was developed to identify the underlying cell types accounted for the observed methylation alterations without the need for cell sorting [29].

At one hand, when one wants to elucidate underlying epigenetic mechanisms of diseases or when the hypothesis is that DNA methylation is a mediator between environmental exposures and human health, it is clear that differences in cell type

composition must be adjusted for. On the other hand, changes in cell type composition may be an important contributor to disease development. For example, a lot of diseases, including metabolic disorders, are accompanied by chronic systemic inflammation and immune cell activation and infiltration. Measuring these cellular heterogeneity may therefore also be useful as prognostic or diagnostic biomarker or as read-out for intervention efficacy. An example is the extrinsic epigenetic clock designed by Horvath, which is influenced by shifts in cell type composition during aging and is able to predict all-cause mortality better than the intrinsic epigenetic clock which is independent of cellular effects [30]. In addition, the extrinsic clock was also found to be much more influenced by different lifestyle factors [31]. Another example, is the prognostic value of neutrophil/lymphocyte ratio (NLR) in several cancers. Recent studies demonstrated the use of blood DNA methylation to measure NLR which has shown its prognostic value in multiple cancers [32]. These data showed that DNA methylation may be a very sensitive way of measuring subtle changes in cell type composition and cellular activation states. For instance, DNA methylation signatures have been described for activated NK and T-cells [33, 34]. In another study, neutrophil activation could be detected in obesity using gene expression and DNA methylation [35]. Similarly, we identified a blood-based DNA methylation profile which was common in different inflammaging diseases, including atherosclerosis, AD, PD and obesity, which to a large extent reflects changes in immune cell type composition (Chapter 6). However, whether this profile can be used as a prognostic marker or used to monitor lifestyle interventions should be further investigated.

In conclusion, accounting for cellular heterogeneity is needed when one wants to investigate causal epigenetic mechanistic pathways or identify new therapeutic targets. Even, when the phenotype of interest doesn't have any effect on cell type heterogeneity, one must take into account the existence of subtle changes in cell type composition and meta-epigenomes when interpreting these data. For biomarker research, correcting for cellular heterogeneity is not an absolute requirement, and can even be exploited as potential prognostic biomarker or to monitor therapeutic or lifestyle interventions.

How to interpret biological impact of DNA methylation changes?

Cause or consequence?

A limitation of the studies performed in this thesis is the observational nature of the study design. For example, in chapter 3 we measured DNA methylation in the children when they were eight years of age, and found an association with prenatal pesticide exposure in interaction with a PON1 genotype. However, we don't know whether these methylation changes were already present at birth and whether DNA methylation is indeed a mediator between prenatal pesticide exposure and the increased cardiometabolic risk profile observed in these children. The cross-sectional case/control design can only demonstrate associations and correlations but doesn't tell us something about the causal relationship. For example, the DNA methylation changes seen in these children can also be the result of the altered lipid levels (which is known as reverse causation). Similarly, in chapter 4 institutionalized children were associated with an hypomethylated profile, but whether this is due to the institutional care or due to some other confounder is not clear and cannot be inferred from this study. In chapter 5, BRCA1 and CRISP2 were found to be differentially methylated in blood samples of atherosclerotic patients, but again we don't know whether these changes were already present before disease onset, and thus represent an epigenetic risk marker, or whether these changes were rather a consequence of the disease process.

This limitation applies to most human epigenetic association studies and is inherent to cross-sectional study designs. Knowledge about the causative relationship between exposure, DNA methylation and phenotype is important when the hypothesis of the study is to understand the mechanism of disease or phenotype. Although longitudinal studies are still scare, they would be helpful in identifying DNA methylation changes preceding the development of the (disease) phenotype and give information about the persistence of DNA methylation changes. DNA methylation changes associated with birth weight and gestational age in cord blood didn't persist at 7 and 17 years [36]. In contrast, another study found that some of the changes at birth could be replicated in mid-childhood, however, this was not true for all the CpG sites [37]. Persistent DNA methylation changes have subsequently also be found to be associated with maternal CRH [38], maternal smoking [39] and prenatal mercury exposure [40]. In general, based on these limited data we can say that some DNA methylation changes identified at birth may persist towards childhood and adolescence, while other CpG sites are reversible and disappear. More longitudinal studies are needed to demonstrate persistent epigenetic effects until adulthood, and whether changes precede disease onset.

Some statistical approaches exist which may help in dissecting causal relationships and have been applied in multiple EWAS. One method is a mediation analysis which uses linear regression to asses causal relationships [41]. The most popular method is the one described by Baron and Kenny [42]. In chapter 3 we used this test to demonstrate the

mediation between prenatal exposure on leptin and body fat accumulation through DNA methylation. Also in other studies, this or alternative mediation methods have been used to prove the mediator role of DNA methylation in linking an exposure with a phenotype or disease [43-46]. This method, however, relies on a few strong assumptions [41]. For example, no unmeasured confounders responsible for the association and no measurement errors may exist. These assumptions are often hard to meet and therefore can result in trivial conclusions. Cross-sectional results should therefore always be interpreted with caution and should be rather seen as indicative for a role of DNA methylation as mediator. Despite some limitations of mediation analysis which is sensitive to unobserved confounders, it remains a valuable method to rank and select interesting associations for further investigation and with the addition of other functional data may help in interpreting EWAS.

Another method is Mendelian randomization which makes use of genetic information to infer causality [47]. Genetic associations are by definition causal (or in linkage disequilibrium with the causal variant) because the DNA sequence remains the same during the lifespan (except for somatic mutations in cancer for example). Therefore reverse causality (=disease causes molecular change) can be excluded in this case. DNA methylation on the other hand is dynamic and can therefore also being influenced by the disease. Mendelian randomization methods use this known causal effect of genetic variants to infer causality. Because alleles are randomly segregated during gamete formation, the association between the genetic variant and the outcome of interest is usually not confounded. This method therefore helps in ruling out confounding effects and reverse causality. This approach was successfully applied in a recent EWAS in relation with blood lipid levels [48]. The authors of the study could show that blood lipids influence DNA methylation in blood cells and not the other way around. Another study showed that DNA methylation alterations were a consequence of obesity rather than a cause [49]. DNA methylation may also cause disease which was recently demonstrated for NAFLD [50]. In another study, using a two-step Mendelian randomization approach it was shown that the effect of smoking on inflammation was mediated by DNA methylation [51]. Because SNPs tend to explain only a small part of the phenotype, in general large sample sizes are needed to obtain enough power to reach significance. Furthermore, not for all phenotypes reliable genetic variants exist and although less strict as compared to mediation analysis, Mendelian randomization does rely on some assumptions of which the validity is not always easy to prove [52].

Despite several limitations, these statistical methods are useful ways to increase the interpretability of cross-sectional EWAS, especially in studies seeking underlying molecular mechanisms of disease or linking environmental exposures with phenotypes. These methods should be preferable supported by other independent approaches,

such as functional assays (CRISPR-Cas based epigenetic editing), animal studies, and longitudinal studies to undisputable prove causality and functionality of a specific epigenetic marker. This does not mean that non-causal DNA methylation signatures are not useful. For example, DNA methylation was shown to be rather a consequence of obesity, but still could predict future diabetes type 2 development [49], and thus holding promise as a marker for risk stratification and personalized medicine. If the study objective is simply to use DNA methylation as prognostic biomarkers or to monitor intervention efficacy, direct evidence of a causal role in disease etiology may not be necessary.

Functional relevance

DNA methylation changes identified in this thesis and in similar studies were relatively small (usually around 5 to 10%), as compared to cancer specific epigenetic changes (frequently > 20%) , and therefore difficult to interpret its biological impact. For instance, a difference of 5% in DNA methylation means than only a small fraction of the cell population in the samples undergoes a methylation change. How to interpret these small changes? Are they functionally relevant? Is there a minimum threshold to elicit a biological relevant response?

A first obvious way to estimate biological significance of weak DNA methylation changes is by mapping the DMPs or DMRs to gene or chromosome regions and to perform pathway analysis. Enriched pathways may reveal information of potential reprogrammed signaling pathways and generate new hypotheses. A first problem with this approach is the mapping of the CpG probes to genes. Especially intergenic CpG sites which are located in regulatory enhancer regions or repeat sequences may suffer from accuracy, since enhancers can form loops interacting and regulating distally located genes. In our analyses, we just removed them from our pathway analysis or mapped the CpG probe to their nearest gene, acknowledging that we may lose some information using this approach. Tools mapping regions to genes, such as GREAT which was originally designed for Chip-Seq data [53], may be used to better annotate DMPs, however improvements are necessary. Another problem with pathway analysis for DNA methylation data is that genes containing more CpG probes are more likely to be selected and may lead to biased enriched pathways. Recent statistical methods were developed [54, 55] to correct for this bias and implementation of these methods may improve interpretability in future EWAS and ranking probes.

Potential functional CpG marks are believed to have an effect on gene expression, and supplementing methylation data with other -omics data would substantially improve the interpretation of EWAS results. However, from this thesis and other studies it

became clear that the relationship between DNA methylation and gene expression is complex, and various DNA methylation changes do not seem to directly influence gene expression. It could be for example that the functional consequences of DNA methylation changes only appear upon secondary stimulation in certain environmental contexts, or that other roles of DNA methylation (alternative splicing, repeat repression, genomic stability, regulation of topologically associated chromatin domains, etc.) may be affected. In future experiments, epigenetic editing tools will be useful to demonstrate causal links between DNA specific methylation changes and corresponding gene expression [56].

Another important aspect of DNA methylation is the sequence context of the change. We know that TFs have an important role in establishing DNA methylation patterns, and that DNA methylation, on the other hand, may influence TF binding [57]. A DNA methylation change in a TF binding site may therefore point to an alteration of TF binding. For example, ATAC-seq could be used in parallel with DNA methylation to identify open accessible chromatin regions and possible regulatory regions in the cell type of interest [58]. It can be expected that reference chromatin maps of different cell types, both in healthy and disease states, may help future EWAS interpretation. Because of the relatively low cost and high reproducibility, we and also other EWAS used the Illumina 450K methylation arrays to measure genome-wide DNA methylation. However, an issue with these arrays is that they cover mainly CGIs and promoter regions, while distally regulatory regions, including enhancers and insulators, are much less well covered, and therefore it can be expected that a lot of information may be lost. Indeed, EWAS in chronic diseases and environmental exposures (see various chapters in this thesis) indicate that especially these regulatory regions are the most variable and probably more important in regulating transcription dynamics. The recently launched EPIC 850k Illumina array (850,000 CpG probes) represents an improved coverage of regulatory regions, but still only covers 3% of all CpGs in the human genome, potentially missing various regions of interest [59]. Whole-genome bisulfite sequencing is the gold standard for DNA methylation studies, but the high cost hampers its use as a high-throughput method in EWAS.

A small DNA methylation change indicates that only a subfraction of the cells display a change in methylation, while the methylation status in other cells remains constant. Is it really a change that we detect, or does the treatment/exposure/disease promotes proliferation or survival of a particular cell subpopulation resulting in an observable methylation change? Even in a homogenous cell culture, some CpG sites have methylation values between 20 and 80% indicating that a fraction of the cells are methylated at that site and another fraction are unmethylated at that site. This indicates that even the homogenous cell line samples reveal a spectrum of epigenomes.

So, a small methylation change may either be the result of a real *de novo* DNA methylation change in a cell, or else by selecting a cell type specific DNA methylome, which shifts the cell population composition. In this respect, a recent study examining the effect of prenatal famine during the Dutch hunger winter on the offspring's DNA methylome, showed that differential survival of embryos with particular methylome profiles were selected and that this mainly explains the observed DNA methylation changes seen in these children [60]. The recent availability of new single cell epigenome sequencing approaches may soon reveal dynamics of epigenetic switches at the single cell level [61].

In conclusion, current EWAS, and small-scaled cohort studies in this thesis, demonstrate that DNA methylation changes are associated with several prenatal and postnatal environmental exposures, and common diseases and complex phenotypes. Now, it is time to move from an association-driven EWAS to a more functional interpretable EWAS by combining DNA methylation data with other genomic data and validation of causal relationships using molecular tests. This will allow more solid interpretation of epigenetic studies and ultimately results in more robust biomarkers and therapy targets.

DNA methylation to monitor health status?

Today many commercial genetic tests exist which use a saliva sample to profile your DNA and give information about your health and disease risks and ancestry. However, as we have seen genetics is only a small part of the story, as lifestyle and environmental factors are playing important roles too, often interacting with genetic factors. Therefore profiling your epigenome may provide an additional necessary layer of information which could be implemented in personalized health tests. Indeed, EWAS showed the plasticity of epigenetic markers and their association with chronic diseases and lifestyle factors. The challenges described in the previous sections, such as tissue-of-interest, causality, and interpretability, should be first addressed to fully exploit the potential of epigenetics in healthcare applications. For example, if we know that an epimutation is present early in disease development and may drive disease progression, this would be an interesting marker for disease risk prediction and may even be a target for therapeutic interventions. Despite these challenges, epigenetic profiling may become a promising biomarker field in health monitoring and disease risk prediction. Whereas genetic factors are static, epigenetic factors are dynamic and are

supposed to be reversible in response to lifestyle interventions. Epigenetic profiling is therefore highly attractive to examine reversible effects of lifestyle interventions and to monitor health status. But are we ready to implement epigenetic markers for medical and health recommendations?

A promising epigenetic marker is the recently developed multi-tissue DNA methylation-based age predictor [62]. Interestingly, an accelerated epigenetic age is associated among others with all-cause mortality, BMI, neurodegenerative disorders, traumatic stress, and therefore represents a measure for biological age [63]. This first generation epigenetic clock has been recently improved by the implementation of clinical measures of phenotypic age predicting life- and health span [64, 65]. Although this may suggest that DNA methylation can be used to inform health status, phenotypic age based on clinical measures still outperforms DNA methylation age in predicting mortality and morbidity [66]. On the other hand, DNA methylation age might be used to predict age acceleration in specific tissues as opposed to clinical measures. It can be expected that tissues may age differently and that this might have health implications. More and more healthtech companies are now providing commercial epigenetic test to predict your biological age. An example is Chronomics which uses a saliva sample to measure your biological age and to give recommendations how to slow down your epigenetic clock. Another company EpigenCare uses DNA methylation to analyze the "age" and quality of your skin and give personalized advise which skincare products and brands you should use. Although studies could associate healthy lifestyle with a slower epigenetic age acceleration, it is not yet known whether relative short term lifestyle interventions are sufficient to slow down epigenetic aging and whether this correlates with a better health. Moreover, it is not clear which algorithms and training datasets have been applied for the health recommendations. Therefore more studies should be performed on the dynamics of epigenetic age before it can be reliable being used as a health marker.

Is there scientific evidence that we can monitor lifestyle interventions in healthy individuals and patients with cardiometabolic disorders? Nutrition has been shown to have a strong impact on DNA methylation *in vitro* studies in animal studies, and human epidemiological studies [67]. However, prolonged human nutritional intervention studies are too limited and for today, no robust diet specific effects on human epigenome have been demonstrated in longitudinal *in vivo* studies. A systematic review on DNA methylation effects of human diet intervention studies showed a global increase in DNA methylation of colorectal mucosa after folic acid supplementation, but not in blood when combining all datasets [68]. Interestingly, genetic factors may influence the DNA methylation response of dietary compounds. For instance, only individuals with a certain polymorphism in the methylenetetrahydrofolate reductase

(MTHFR) gene demonstrated a change in methylation in multiple studies [68]. MTHFR plays an important role in the formation of SAM, the substrate for DNMT, and changes in activity of this enzyme are expected to influence DNA methylation. Another systematic review, where the effect of nutritional interventions during pregnancy on offspring's DNA methylation was examined, showed little evidence for significant DNA methylation changes in humans [69]. However, the offspring's DNA methylation responsiveness towards maternal nutritional interventions is stronger when subgrouping the mothers by BMI, smoking and nutritional status. This indicates that beside genetic factors, also lifestyle factors may interact with maternal nutritional supplementation [69]. Beside diet intervention studies, also other lifestyle interventions such as exercise seems to interfere with DNA methylation [70, 71], but also here studies are rather small and should be seen as explorative pilot results.

It is well known that every individual responds different to lifestyle factors. Can we use DNA methylation to predict dietary health effects? Can we use DNA methylation for personalized nutrition? So far, the evidence is limited. Response to weight-loss interventions have been associated with DNA methylation and transcriptomic differences [72, 73]. Another study showed that vitamin D response was dependent on the methylation levels of the Cytochrome P450 enzymes CYP2R1 and CYP24A1 [74]. These preliminary results may hint the importance of DNA methylation in predicting dietary responses, but we are still far from using epigenetic markers in personalized nutrition. The future of personalized nutrition and personalized lifestyle advise will also lie in the integration of different -omics datasets together with other health and clinical measures. For example, a recent study indicates the importance of the gut microbiome in personalized nutrition [75]. Furthermore, the microbiome itself may also influence epigenetic patterns [76]. As we gather more data and combine this with machine learning and artificial intelligence techniques, personalized health advise may be possible in the future. For example, mobile apps and wearables may track real-time health measures and activity. In addition, apps exist which collects all your health data and matches your health measures with other individuals to give more personalized advise in managing particular chronic disorders (see https://doc.ai/ as example). In this respect, it is not unthinkable that also epigenetic information may be implemented in future disease risk predictions and tools to monitor health status.

Conclusion and future directions

The studies in this thesis and in literature identified many associations between DNA methylation and environmental factors or chronic diseases. However, the biological

relevance of these associations are not yet known. Improving the functional interpretation of epigenetic changes is an absolute requirement to implement epigenetic markers in clinical practice, give insight in disease mechanisms and define new epigenetic targets for therapy. Future EWAS should therefore move from association-based studies to interpretation-based studies by combining multiple molecular datasets (RNAseq, ATACseq, etc), and performing functional assays (epigenetic editing) or longitudinal studies.

Future EWAS studies will also benefit from new reference methylomes constructed by international consortia, like BLUEPRINT and IHEC to identify subtle shifts in underlying cell type contribution. This may be further enhanced by improvements in single-cell technologies. Comparing methylation profiles from different tissues of the same individual will allow us to detect tissue-specific methylation changes and changes which are in common across different tissues. This information may be valuable in identifying robust methylation markers in easy-accessible tissue such as blood and saliva.

The most recent methylation array only covers 3% of all CpG sites in the human genome, therefore new methods which are cost-efficient and enhance CpG coverage may improve future EWAS as most variable sites are located between genes or in gene bodies. Furthermore, most of the existing DNA methylation profiling methods depend on bisulfite conversion which is a harsh treatment and is not able to discriminate between 5mC and 5hmC. Therefore, the development of bisulfite-free sequencing methods (SMRT, nanopore sequencing) may open new perspectives. More studies are now also elucidating the importance of non-CpG methylation, other DNA modifications, such as N6-methadenine (6mA), and even RNA modifications. More insight in the function of these epigenetic modifications may open new opportunities for epigenetic biomarker research.

Despite these challenges, it is clear that future applications in personalized medicine and health will benefit from the implementation of epigenetic markers.

References

1. Huang, Y.T., et al., *Epigenome-wide profiling of DNA methylation in paired samples of adipose tissue and blood.* Epigenetics, 2016. **11**(3): p. 227-36.
2. Dick, K.J., et al., *DNA methylation and body-mass index: a genome-wide analysis.* Lancet, 2014. **383**(9933): p. 1990-8.
3. Pfeiffer, L., et al., *DNA methylation of lipid-related genes affects blood lipid levels.* Circ Cardiovasc Genet, 2015. **8**(2): p. 334-42.
4. Houde, A.A., et al., *Leptin and adiponectin DNA methylation levels in adipose tissues and blood cells are associated with BMI, waist girth and LDL-cholesterol levels in severely obese men and women.* BMC Med Genet, 2015. **16**: p. 29.
5. Crujeiras, A.B., et al., *DNA methylation map in circulating leukocytes mirrors subcutaneous adipose tissue methylation pattern: a genome-wide analysis from non-obese and obese patients.* Sci Rep, 2017. **7**: p. 41903.
6. Hannon, E., et al., *Interindividual methylomic variation across blood, cortex, and cerebellum: implications for epigenetic studies of neurological and neuropsychiatric phenotypes.* Epigenetics, 2015. **10**(11): p. 1024-32.
7. Edgar, R.D., et al., *BECon: a tool for interpreting DNA methylation findings from blood in the context of brain.* Transl Psychiatry, 2017. **7**(8): p. e1187.
8. Walton, E., et al., *Correspondence of DNA Methylation Between Blood and Brain Tissue and Its Application to Schizophrenia Research.* Schizophr Bull, 2016. **42**(2): p. 406-14.
9. Lunnon, K., et al., *Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease.* Nat Neurosci, 2014. **17**(9): p. 1164-70.
10. Masliah, E., et al., *Distinctive patterns of DNA methylation associated with Parkinson disease: identification of concordant epigenetic changes in brain and peripheral blood leukocytes.* Epigenetics, 2013. **8**(10): p. 1030-8.
11. Wockner, L.F., et al., *Genome-wide DNA methylation analysis of human brain tissue from schizophrenia patients.* Transl Psychiatry, 2014. **4**: p. e339.
12. Wren, M.E., E.A. Shirtcliff, and S.S. Drury, *Not all biofluids are created equal: chewing over salivary diagnostics and the epigenome.* Clin Ther, 2015. **37**(3): p. 529-39.
13. Langie, S.A.S., et al., *Salivary DNA Methylation Profiling: Aspects to Consider for Biomarker Identification.* Basic Clin Pharmacol Toxicol, 2017. **121 Suppl 3**: p. 93-101.
14. Langie, S.A., et al., *Whole-Genome Saliva and Blood DNA Methylation Profiling in Individuals with a Respiratory Allergy.* PLoS One, 2016. **11**(3): p. e0151109.
15. Wu, H.C., et al., *Correlation of DNA methylation levels in blood and saliva DNA in young girls of the LEGACY Girls study.* Epigenetics, 2014. **9**(7): p. 929-33.
16. Thompson, T.M., et al., *Comparison of whole-genome DNA methylation patterns in whole blood, saliva, and lymphoblastoid cell lines.* Behav Genet, 2013. **43**(2): p. 168-76.
17. Smith, A.K., et al., *DNA extracted from saliva for methylation studies of psychiatric traits: evidence tissue specificity and relatedness to brain.* Am J Med Genet B Neuropsychiatr Genet, 2015. **168B**(1): p. 36-44.
18. Lin, X., et al., *Choice of surrogate tissue influences neonatal EWAS findings.* BMC Med, 2017. **15**(1): p. 211.
19. Jiang, R., et al., *Discordance of DNA methylation variance between two accessible human tissues.* Sci Rep, 2015. **5**: p. 8257.
20. Braun, P.R., et al., *Genome-wide DNA methylation comparison between live human brain and peripheral tissues within individuals.* Transl Psychiatry, 2019. **9**(1): p. 47.
21. Liu, Y., et al., *Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis.* Nat Biotechnol, 2013. **31**(2): p. 142-7.
22. Jaffe, A.E. and R.A. Irizarry, *Accounting for cellular heterogeneity is critical in epigenome-wide association studies.* Genome Biol, 2014. **15**(2): p. R31.
23. Houseman, E.A., et al., *DNA methylation arrays as surrogate measures of cell mixture distribution.* BMC Bioinformatics, 2012. **13**: p. 86.

24. Teschendorff, A.E. and S.C. Zheng, *Cell-type deconvolution in epigenome-wide association studies: a review and recommendations.* Epigenomics, 2017. **9**(5): p. 757-768.
25. Bauer, M., et al., *A varying T cell subtype explains apparent tobacco smoking induced single CpG hypomethylation in whole blood.* Clin Epigenetics, 2015. **7**: p. 81.
26. Lappalainen, T. and J.M. Greally, *Associating cellular epigenetic models with human phenotypes.* Nat Rev Genet, 2017. **18**(7): p. 441-451.
27. Houseman, E.A., et al., *DNA Methylation in Whole Blood: Uses and Challenges*. Curr Environ Health Rep, 2015. **2**(2): p. 145-54.
28. Wijetunga, N.A., et al., *The meta-epigenomic structure of purified human stem cell populations is defined at cis-regulatory sequences.* Nat Commun, 2014. **5**: p. 5195.
29. Zheng, S.C., et al., *Identification of differentially methylated cell types in epigenome-wide association studies.* Nat Methods, 2018. **15**(12): p. 1059-1066.
30. Chen, B.H., et al., *DNA methylation-based measures of biological age: meta-analysis predicting time to death.* Aging (Albany NY), 2016. **8**(9): p. 1844-1865.
31. Quach, A., et al., *Epigenetic clock analysis of diet, exercise, education, and lifestyle factors.* Aging (Albany NY), 2017. **9**(2): p. 419-446.
32. Koestler, D.C., et al., *DNA Methylation-Derived Neutrophil-to-Lymphocyte Ratio: An Epigenetic Tool to Explore Cancer Inflammation and Outcomes.* Cancer Epidemiol Biomarkers Prev, 2017. **26**(3): p. 328-338.
33. Komori, H.K., et al., *Defining CD4 T cell memory by the epigenetic landscape of CpG DNA methylation.* J Immunol, 2015. **194**(4): p. 1565-79.
34. Wiencke, J.K., et al., *The DNA methylation profile of activated human natural killer cells.* Epigenetics, 2016. **11**(5): p. 363-80.
35. Pan, Y., et al., *Discovery and Validation of a Novel Neutrophil Activation Marker Associated with Obesity.* Sci Rep, 2019. **9**(1): p. 3433.
36. Simpkin, A.J., et al., *Longitudinal analysis of DNA methylation associated with birth weight and gestational age.* Hum Mol Genet, 2015. **24**(13): p. 3752-63.
37. Agha, G., et al., *Birth weight-for-gestational age is associated with DNA methylation at birth and in childhood.* Clin Epigenetics, 2016. **8**: p. 118.
38. Tian, F.Y., et al., *Maternal corticotropin-releasing hormone is associated with LEP DNA methylation at birth and in childhood: an epigenome-wide study in Project Viva.* Int J Obes (Lond), 2018.
39. Richmond, R.C., et al., *Prenatal exposure to maternal smoking and offspring DNA methylation across the lifecourse: findings from the Avon Longitudinal Study of Parents and Children (ALSPAC).* Hum Mol Genet, 2015. **24**(8): p. 2201-17.
40. Cardenas, A., et al., *Persistent DNA methylation changes associated with prenatal mercury exposure and cognitive performance during childhood.* Sci Rep, 2017. **7**(1): p. 288.
41. Richmond, R.C., et al., *Challenges and novel approaches for investigating molecular mediation.* Hum Mol Genet, 2016. **25**(R2): p. R149-R156.
42. Baron, R.M. and D.A. Kenny, *The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations.* J Pers Soc Psychol, 1986. **51**(6): p. 1173-82.
43. Houtepen, L.C., et al., *Genome-wide DNA methylation levels and altered cortisol stress reactivity following childhood trauma in humans.* Nat Commun, 2016. **7**: p. 10967.
44. Tobi, E.W., et al., *DNA methylation as a mediator of the association between prenatal adversity and risk factors for metabolic disease in adulthood.* Sci Adv, 2018. **4**(1): p. eaao4364.
45. Kupers, L.K., et al., *DNA methylation mediates the effect of maternal smoking during pregnancy on birthweight of the offspring.* Int J Epidemiol, 2015. **44**(4): p. 1224-37.
46. Cao-Lei, L., et al., *DNA methylation mediates the impact of exposure to prenatal maternal stress on BMI and central adiposity in children at age 13(1/2) years: Project Ice Storm.* Epigenetics, 2015. **10**(8): p. 749-61.
47. Davey Smith, G. and G. Hemani, *Mendelian randomization: genetic anchors for causal inference in epidemiological studies.* Hum Mol Genet, 2014. **23**(R1): p. R89-98.

48. Dekkers, K.F., et al., *Blood lipids influence DNA methylation in circulating cells.* Genome Biol, 2016. **17**(1): p. 138.
49. Wahl, S., et al., *Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity.* Nature, 2017. **541**(7635): p. 81-86.
50. Ma, J., et al., *A Peripheral Blood DNA Methylation Signature of Hepatic Fat Reveals a Potential Causal Pathway for Non-Alcoholic Fatty Liver Disease.* Diabetes, 2019.
51. Jhun, M.A., et al., *Modeling the Causal Role of DNA Methylation in the Association Between Cigarette Smoking and Inflammation in African Americans: A 2-Step Epigenetic Mendelian Randomization Study.* Am J Epidemiol, 2017. **186**(10): p. 1149-1158.
52. Latvala, A. and M. Ollikainen, *Mendelian randomization in (epi)genetic epidemiology: an effective tool to be handled with care.* Genome Biol, 2016. **17**(1): p. 156.
53. McLean, C.Y., et al., *GREAT improves functional interpretation of cis-regulatory regions.* Nat Biotechnol, 2010. **28**(5): p. 495-501.
54. Geeleher, P., et al., *Gene-set analysis is severely biased when applied to genome-wide methylation data.* Bioinformatics, 2013. **29**(15): p. 1851-7.
55. Dong, D., et al., *ebGSEA: An improved Gene Set Enrichment Analysis method for Epigenome-Wide-Association Studies.* Bioinformatics, 2019.
56. Kungulovski, G. and A. Jeltsch, *Epigenome Editing: State of the Art, Concepts, and Perspectives.* Trends Genet, 2016. **32**(2): p. 101-113.
57. Zhu, H., G. Wang, and J. Qian, *Transcription factors as readers and effectors of DNA methylation.* Nat Rev Genet, 2016. **17**(9): p. 551-65.
58. Buenrostro, J.D., et al., *Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position.* Nat Methods, 2013. **10**(12): p. 1213-8.
59. Pidsley, R., et al., *Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling.* Genome Biol, 2016. **17**(1): p. 208.
60. Tobi, E.W., et al., *Selective Survival of Embryos Can Explain DNA Methylation Signatures of Adverse Prenatal Environments.* Cell Rep, 2018. **25**(10): p. 2660-2667 e4.
61. Hui, T., et al., *High-Resolution Single-Cell DNA Methylation Measurements Reveal Epigenetically Distinct Hematopoietic Stem Cell Subpopulations.* Stem Cell Reports, 2018. **11**(2): p. 578-592.
62. Horvath, S., *DNA methylation age of human tissues and cell types.* Genome Biol, 2013. **14**(10): p. R115.
63. Declerck, K. and W. Vanden Berghe, *Back to the future: Epigenetic clock plasticity towards healthy aging.* Mech Ageing Dev, 2018. **174**: p. 18-29.
64. Levine, M.E., et al., *An epigenetic biomarker of aging for lifespan and healthspan.* Aging (Albany NY), 2018. **10**(4): p. 573-591.
65. Lu, A.T., et al., *DNA methylation GrimAge strongly predicts lifespan and healthspan.* Aging (Albany NY), 2019. **11**(2): p. 303-327.
66. Horvath, S. and K. Raj, *DNA methylation-based biomarkers and the epigenetic clock theory of ageing.* Nat Rev Genet, 2018. **19**(6): p. 371-384.
67. Szarc vel Szic, K., et al., *From inflammaging to healthy aging by dietary lifestyle choices: is epigenetics the key to personalized nutrition?* Clin Epigenetics, 2015. **7**: p. 33.
68. ElGendy, K., et al., *Effects of dietary interventions on DNA methylation in adult humans: systematic review and meta-analysis.* Br J Nutr, 2018. **120**(9): p. 961-976.
69. Andraos, S., et al., *The Impact of Nutritional Interventions in Pregnant Women on DNA Methylation Patterns of the Offspring: A Systematic Review.* Mol Nutr Food Res, 2018. **62**(24): p. e1800034.
70. Ronn, T., et al., *A six months exercise intervention influences the genome-wide DNA methylation pattern in human adipose tissue.* PLoS Genet, 2013. **9**(6): p. e1003572.
71. Barres, R., et al., *Acute exercise remodels promoter methylation in human skeletal muscle.* Cell Metab, 2012. **15**(3): p. 405-11.
72. Bouchard, L., et al., *Differential epigenomic and transcriptomic responses in subcutaneous adipose tissue between low and high responders to caloric restriction.* Am J Clin Nutr, 2010. **91**(2): p. 309-20.

73. Moleres, A., 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, 2013. **27**(6): p. 2504-12.
74. Zhou, Y., et al., *DNA methylation levels of CYP2R1 and CYP24A1 predict vitamin D response variation.* J Steroid Biochem Mol Biol, 2014. **144 Pt A**: p. 207-14.
75. Zeevi, D., et al., *Personalized Nutrition by Prediction of Glycemic Responses.* Cell, 2015. **163**(5): p. 1079-1094.
76. Qin, Y. and P.A. Wade, *Crosstalk between the microbiome and epigenome: messages from bugs.* J Biochem, 2018. **163**(2): p. 105-112.

Acknowlegdements

‘Dat wordt dan negen jaar.’ Ik hoor het mijn moeder nog zo zeggen. We staan aan de infostand Biochemie & Biotechnologie, op een aangename lentedag in april 2008. ‘Ja dat zal wel’ dacht ik bij mezelf. ‘Eerst maar eens dat eerste jaar overleven en dan zien we wel’ starend naar de centimeters dikke boeken en cursussen die voor mij lagen. Uiteindelijk werden het 11 jaar, vijf jaar bachelor en master en zes jaar doctoraat.

Het zijn zes ontzettend boeiende jaren geweest. Uiteraard waren er twijfels en onzekerheden, maar uiteindelijk zorgen die er net voor dat de voldoening achteraf groter is. Onzekerheden zorgen voor groei. Als je alles denkt te weten maak je geen vorderingen. Dit doctoraat heeft me niet enkel wetenschappelijke kennis en vaardigheden opgebracht, maar heeft er ook voor gezorgd dat ik mezelf nu veel beter ken.

De eerste keer dat ik in het T gebouw kwam was voor een practicum tijdens het 2e bachelor jaar. Ik weet nog dat ik dacht, hier wil ik nooit werken. Veel te warm, veel te donker, het constante gezoem van diepvriezers, behangpapier die van de muren kwam, ontzettend lelijke oranje deuren en een laag plafond die het nog meer claustrofobisch en deprimerend maakte. Nee, hier zou ik nooit willen werken. De werkomgeving is ook belangrijk!
Drie jaar later sta ik voor zo’n zelfde lelijke oranje deur met een registratieformulier in de hand om bij Prof. Wim vanden Berghe te komen werken als thesis student. Het thesisonderwerp heeft het gehaald van de werksfeer. Gedachten kunnen veranderen.
Ik had geluk. Wim was aanwezig. Later zou ik ontdekken dat dit eerder uitzonderlijk was. Voor mij was het begin van een bijna zeven jaar durende samenwerking.
Wim, ik heb je leren kennen als de meest open-gesloten persoon die ik ken (wellicht ben ik de meest gesloten-gesloten persoon, maar dit terzijde). Altijd de zotste verhalen, terecht komend in de gekste situaties die niemand anders kan verzinnen, maar uiteindelijk komt altijd alles goed. Echt waar, uiteindelijk komt alles goed. Ongelooflijk. Maar op andere momenten was het veel moeilijker om je te begrijpen: voor drie weken naar Cuba vertrekken zonder ons in te lichten, studenten die voor onze deur staan om te starten met hun thesis/stage zonder dat we iets weten (‘oei, en Wim is er vandaag uitzonderlijk niet…’), nog last-minute iemand zoeken om examentoezicht te geven (daar gaat mijn planning), en ga zo maar door... Voor iemand die structuur en orde nodig heeft, was het soms wat chaos.

Ik had het zelf allemaal meegemaakt tijdens mijn eerste dag thesis. Rond 9 uur zal er wel iemand zijn dacht ik. Niemand te bespeuren. Uiteindelijk toch iemand gevonden (Kasia, zou later blijken). Met verbaasde blik 'no, I didn't know that you will start today. I'm not sure if Wim will come to the lab today'. Ik voelde me meteen welkom (of niet). Ondanks alles (viel wel mee), heb ik een enorme bewondering voor je positieve ingesteldheid en passie voor wetenschap. Je kennis is enorm. Ik kon me echt soms dom voelen tijdens meetings waarbij je de ene interessante studie na de andere opsomde. Een levende encyclopedie. Ik heb ontzettend veel bijgeleerd en mogen bijleren, en daarvoor wil ik je ontzettend hard bedanken. Dankzij jou heb ik de kans gekregen om dingen te doen die ik anders nooit zou durven, zoals workshops en cursussen geven in het buitenland, een talk geven voor een volle zaal experten, etc. Jij hebt er mee voor gezorgd dat ik gegroeid ben zowel als wetenschappers en als mens. Echt waar, ik heb er van genoten. Bedankt!

De meeste projecten in dit doctoraat waren uiteraard nooit mogelijk geweest zonder de hulp van anderen. Ik zou graag Sylvie Rémy willen bedanken om mij te helpen bij de statistische verwerking tijdens het PON1 project. Jij hebt me ook de extra stimulans gegeven om mezelf te verdiepen in statistische data analyse. Ik wil hierbij ook Greet Schoeters en Helle Andersen bedanken die mij de kans gaven om op dit enorm interessant project te werken, en de resultaten mee te interpreteren en bediscussiëren. Ik zou ook graag Patrick De Boever en Stacy Drury willen bedanken om mij mee te betrekken in het Bucharest Early Intervention Project, en om samen inzicht te krijgen in de resultaten. Ik hoop echt dat we dit project nog kunnen publiceren. Geoffrey Istas wil ik bedanken voor de prettige samenwerking tijdens het atherosclerose project. Dragan Milenkovic bedank ik voor de samenwerking, inzichten en discussies tijdens de flaviola studies.

Tijdens mijn doctoraat heb ik het geluk gehad om met verschillende interessante mensen te mogen samenwerken met verschillende achtergronden en persoonlijkheden. Allemaal hebben jullie rechtstreeks of onrechtstreeks invloed gehad op mij als persoon en hoe dit doctoraat tot stand is gekomen. De werkomgeving is voor mij heel belangrijk. Ik heb een plaats nodig waar ik op mijn gemak ben, waar ik zonder veel stress of spanningen kan werken, dan presteer ik het best. Het PPES labo was zo'n plaats. Ik wil dan ook al de mensen bedanken die mee hebben bijgedragen aan de aangename sfeer in het labo: Ruben, Martin, Mops, Geert, Daisy, Dietmar, Joey, Zainab, Nils, Herald, Marie-Louise, Evi, Zoë, An, Stijn, Joke, Wendy, Sylvia, Xaveer. Bedankt!
Kasia, bedankt voor mij in te leiden in de wondere wereld van de epigenetica en DNA methylatie. Als senior wetenschapper gaf je me tal van nuttige tips hoe experimenten

op te zetten en hoe wetenschappelijk teksten te schrijven. Ondanks je – het moet gezegd - nogal ongeduldig karakter, heb je toch al mijn vele vragen (zeker in het begin) willen beantwoorden. Chandra en Ajay, de Indiërs in ons labo, ook bij jullie kon ik steeds terecht bij vragen of bedenkingen. Wat ik vooral bewonder aan jullie is hoe onvoorwaardelijk Indiërs elkaar steunen en hoe open jullie zijn ten opzichte van anderen. Jolien, ik had het gevoel dat jij met iedereen wel overweg kon. Jij was echt de spreekwoordelijke 'lijm' die de groep bijeen hield en initiatief nam om tal van groepsactiviteiten te organiseren. Je was een heel aangename persoon om mee samen te werken. Ik heb ook het geluk gehad om een tijdje met Farnaz te kunnen samenwerken. Farnaz, in veel opzichten zijn wij tegenpolen, ik de rust zelve, jij de spontane (en soms wel wat over-paniekerige). Maar op één of andere manier werkte dit heel goed. Jij kon bepaalde zaken gedaan krijgen waarvoor ik niet assertief genoeg was. Je wordt nog altijd gemist op het labo. Op sommige momenten denk ik echt, nu zouden we Farnaz nodig hebben. Claudina, ik heb echt een enorme bewondering voor jou. Ik denk dat wij elkaar de eerste keer ontmoet hebben in UGent voor de start van het echinacea project. Toen al viel op hoe behulpzaam je was en bent, en hoe je ervoor zorgt dat iedereen zich welkom voelt. Emilie, ook jij was nooit te beroerd om andere mensen te helpen en ervoor te zorgen dat elke nieuwkomer de juiste begeleiding kreeg. Ik bewonder echt de manier hoe jij ondanks al de moeilijkheden tijdens het begin van je doctoraat, blijft doorzetten. Hoe jij de verantwoordelijkheid, initiatief en leiding neemt in het labo is bewonderingswaardig. Martijn, de enige die mijn interesse voor wielrennen en veldrijden deelde. Bedankt ook voor de gezellige vrijdagnamiddag borrels! Ruben, elke dag hadden we een afspraak om 12u, lunchtime! Ik denk dat we er samen niet veel gemist hebben. Je flauwe doch grappige mopjes zorgden voor een enorm aangename werksfeer.
Speciale dank ook aan al de Cubanen die ik heb mogen ontmoeten: Laura, Ivones, Carlos, Elisabeth R., Elisabeth, Julio, Alejandro, René, Idania, José, Gilberto, Olivia en wellicht ben ik er nog een aantal vergeten. Soms leek het wel een Cubaanse invasie. Gelukkig kwamen jullie in vrede en was het vaak een plezier om jullie bezig te zien. Ik werd daar gelukkig van (met mate).

Ik heb ook de kans gekregen om tijdens mijn doctoraat practicums mee te assisteren. Eva, Karen, Manou en Stephanie, bedankt voor de leuke tijd tijdens het practicum. Ik had tijdens mijn studies nooit gedacht dat ik zelf ooit practicum zou geven. Ondanks ik dit echt vermoeiend vond (s' avonds was ik echt op van een teveel aan prikkels), moet ik zeggen dat ik het toch wel aangenaam vond om andere studenten mee te kunnen helpen. Het bracht mij terug naar mijn studententijd; nog totaal onwetend en onzeker over wat de toekomst zou brengen.

Het laatste woord gaat uit naar mijn familie, wellicht de belangrijkste mensen in mijn leven. Mijn drieling-zus Carmen, ik kan echt jaloers zijn op jouw enthousiasme. De manier waarop jij ergens invliegt en er telkens volledig (vaak over het randje) ervoor gaat is bewonderingswaardig. Mijn jongste broer Jim, ik ben echt jaloers op jouw durf en zelfvertrouwen. Een beetje tegen onze familietraditie in ben jij de extravertste van ons gezin, iemand die zaken gedaan kan krijgen. Ik wou dat ik het ook kon. Mijn drieling-broer Glenn, wij hebben meestal niet veel woorden nodig om elkaar te begrijpen. Wellicht ben jij de enigste die mij volledig kan doorgronden. Wij hebben lange tijd alles samen gedaan en het was dan ook aanpassen wanneer we na het middelbaar onze eigen weg volgden. Ik heb nog lange tijd in de wij-vorm gesproken in plaats van de ik-vorm. Mijn papa, wellicht de rustigste persoon die ik ken. Hoe jij schijnbaar stress-loos in het leven staat is prachtig. Jij hebt me geleerd om vooral je eigen ding te doen en vooral niet de grote stroom te volgen. Weg met de conventies. Mijn mama, bij problemen kom ik altijd eerst bij u terecht, dat is waar. Maar jij bent dan ook diegene die mij telkens opnieuw gerust kan stellen en mij onvoorwaardelijk steunt. Dank om mij te kalmeren wanneer ik weer eens het verkeerde examen heb ingestudeerd, om mij last-minute nog te vervoeren naar de campus omdat de treinen op examen-dag besluiten niet te rijden, om mij om middernacht nog op te pikken na een congres toen ik vastzat in de luchthaven, om mijn diploma en prijs te ontvangen omdat ik persé nog op scoutskamp moest, etc. Het kan soms eng aanvoelen wanneer je beseft dat je zelf de verantwoordelijkheid draagt over je leven. Het is dan ontzettend aangenaam om te weten dat je bij twijfel terug kan vallen op je familie. Bedankt voor alles!

Academic Curriculum Vitae

Ken Declerck

Oude-Godstraat 71
2650 Edegem, Belgium
+32497648617
kendeclerck90@hotmail.com
ken.declerck@uantwerpen.be
Born in Edegem, 7 December 1990

Work experience

2013-2019	**PhD Student** Laboratory of Protein Chemistry, Proteomics & Epigenetic signalling (PPES) Department of Biomedical Sciences University of Antwerp Supervisor: Prof. Dr. Wim Vanden Berghe Project title: "*Epigenetic profiling of adverse lifestyle conditions and nutraceutical interventions in health and disease*"
2014-2019	**Academic assistant** Assisting in practicums 2nd Ba, 3rd Ba and 1st Ma Biomedical Sciences Department of Biomedical Sciences University of Antwerp

Education

2013 **Master of Science (MSc): Biochemistry & Biotechnology: Molecular and Cellular Gene Biotechnology** (*great distinction*)
Department of Biomedical Sciences
University of Antwerp

Master thesis: "*Vergelijking van Illumina 450K CpG array en CpG pyrosequencing gebaseerde DNA-methylatie-analysemethodes in onderzoek naar cardioprotectieve epigenetische effecten van flavanolen*"
Supervisor: Prof. Dr. Wim Vanden Berghe

2011 **Bachelor of Science (BSc): Biochemistry & Biotechnology** (*great distinction*)
Department of Biomedical Sciences
University of Antwerp

Bachelor thesis: "*Distributie van leptine en ghreline in de big tijdens de perinatale periode*"
Supervisor: Prof. Dr. Chris Van Ginneken

2008 **Wetenschappen-wiskunde**
Moretus Ekeren

Teaching experience

2014-2019 **Academic assistant**
Assisting in practicums 2nd Ba, 3rd Ba and 1st Ma Biomedical Sciences
Department of Biomedical Sciences
University of Antwerp

2017 Teacher Training School: "Use of nutrigenetics & nutri(epi)genomics in nutrition research"
POSITIVe COST action
Thessaloniki (Greece)

2015 Teacher one-week workshop “Transcriptomics & Bioinformatics”

VLIR Network Ecuador
Espol, Guayaquil (Ecuador)

Courses & Certificates

2017 **“Analysis of grouped and longitudinal data using linear mixed models” course**
StatUA, University of Antwerp

2015 **“Introduction to Linux for bioinformatics” training**
VIB Bioinformatics Core, BITS

2015 **Workshop “From big-data to bedside”**
Center for Medical Genetics, UGent

2014 **“Introduction to Affymetrix microarray analysis” training**
VIB Bioinformatics Core, BITS

2014 **“Building bioinformatics workflows” training**
VIB Bioinformatics Core, BITS

2014 **R-workshop**
StatUA, University of Antwerp

2013 **Laboratory animal leader (FELASA Cat C)**
University of Antwerp

DataCamp online courses
Completed multiple data science courses on data manipulation, visualization, statistics and machine learning in R

Skills

Computer skills	R programming, basic knowledge of linux, use of bioinformatic tools (pathway enrichment tools, Genome Browsers, CytoScape)
Data analysis skills	Experienced in analyzing -omics datasets (Illumina DNA methylation arrays, gene expression arrays, RNAseq, kinome data), data manipulation & cleaning (tidyverse), data visualization & exploration (ggplot2), unsupervised clustering techniques (hierarchal clustering, principal component analysis), excellent statistical knowledge (linear models, mixed models, mediation analysis)
Lab skills	qPCR, pyrosequencing, basic cell culture, western-blot, ChIP
Languages	Dutch (native language), English (fluent), French (basic)

Publications

Peer-reviewed:

2019:

Van Acker ZP, **Declerck K**, Luyckx E, Vanden Berghe W, Dewilde S. *Non-methylation-Linked Mechanism of REST-Induced Neuroglobin Expression Impacts Mitochondrial Phenotypes in a Mouse Model of Amyotrophic Lateral Sclerosis.* Neuroscience 2019 May 31. pii: S0306-4522(19)30369-0. doi: 3110.1016/j.neuroscience.2019.05.039

2018:

Milenkovic D, Berghe WV, Morand C, Claude S, van de Sandt A, Gorressen S, Monfoulet LE, Chirumamilla CS, **Declerck K**, Szic KSV, Lahtela-Kakkonen M, Gerhauser C, Merx MW, Kelm M. *A systems biology network analysis of nutri(epi)genomic changes in endothelial cells exposed to epicatechin metabolites.* Sci Rep. 2018 Oct 19;8(1):15487. doi: 10.1038/s41598-018-33959-x.

Declerck K, Vanden Berghe W. *Back to the future: Epigenetic clock plasticity towards healthy aging.* Mech Ageing Dev. 2018 Sep;174:18-29. doi: 10.1016/j.mad.2018.01.002.

Hassannia B, Wiernicki B, Ingold I, Qu F, Van Herck S, Tyurina YY, Bayır H, Abhari BA, Angeli JPF, Choi SM, Meul E, Heyninck K, **Declerck K**, Chirumamilla CS, Lahtela-Kakkonen M, Van Camp G, Krysko DV, Ekert PG, Fulda S, De Geest BG, Conrad M, Kagan VE, Vanden Berghe W, Vandenabeele P, Vanden Berghe T. *Nano-targeted induction of dual ferroptotic mechanisms eradicates high-risk neuroblastoma.* J Clin Invest. 2018 Aug 1;128(8):3341-3355. doi: 10.1172/JCI99032.

2017:

Chirumamilla CS, Palagani A, Kamaraj B, **Declerck K**, Verbeek MWC, Oksana R, De Bosscher K, Bougarne N, Ruttens B, Gevaert K, Houtman R, De Vos WH, Joossens J, Van Der Veken P, Augustyns K, Van Ostade X, Bogaerts A, De Winter H, Vanden Berghe W. *Selective Glucocorticoid Receptor Properties of GSK866 Analogs with Cysteine Reactive Warheads.* Front Immunol. 2017 Nov 1;8:1324. doi: 10.3389/fimmu.2017.01324.

García-Pupo L, Sánchez JR, Ratman D, Pérez-Novo C, **Declerck K**, De Bosscher K, Markakis MN, Beemster G, Zaldo A, Nuñez Figueredo Y, Delgado-Hernández R, Vanden Berghe W. *Semi-synthetic sapogenin exerts neuroprotective effects by skewing the brain ischemia reperfusion transcriptome towards inflammatory resolution.* Brain Behav Immun. 2017 Aug;64:103-115. doi: 10.1016/j.bbi.2017.04.002.

Declerck K*, Istas G*, Pudenz M, Szarc vel Szic K, Lendinez-Tortajada V, Leon-Latre M, Heyninck K, Haegeman G, Casasnovas JA, Tellez-Plaza M, Gerhauser C, Heiss C, Rodriquez-Mateos A, Vanden Berghe W. *Identification of differentially methylated BRCA1 and CRISP2 DNA regions as blood surrogate markers for cardiovascular disease*. Sci Rep. 2017 Jul 11;7(1):5120. doi: 10.1038/s41598-017-03434-0 (*Equal contribition)

Szarc Vel Szic K, **Declerck K**, Crans RAJ, Diddens J, Scherf DB, Gerhäuser C, Vanden Berghe W. *Epigenetic silencing of triple negative breast cancer hallmarks by Withaferin A.* Oncotarget. 2017 Jun 20;8(25):40434-40453. doi: 10.18632/oncotarget.17107.

Declerck K, Remy S, Wohlfahrt-Veje C, Main KM, Van Camp G, Schoeters G, Vanden Berghe W, Andersen HR. *Interaction between prenatal pesticide exposure and a*

common polymorphism in the PON1 gene on DNA methylation in genes associated with cardio-metabolic disease risk—an exploratory study. Clin Epigenetics. 2017 Apr 5;9:35. doi: 10.1186/s13148-017-0336-4.

2016:

Langie SA, Moisse M, **Declerck K**, Koppen G, Godderis L, Vanden Berghe W, Drury S, De Boever P. *Salivary DNA methylation profiling: Aspects to consider for biomarker identification.* Basic Clin Pharmacol Toxicol. 2016 Nov 30. doi: 10.1111/bcpt.12721. Review.

Langie SA, Szarc Vel Szic K, **Declerck K**, Traen S, Koppen G, Van Camp G, Schoeters G, Vanden Berghe W, De Boever P. *Whole-genome saliva and blood DNA methylation profiling in individuals with a respiratory allergy.* PLoS One. 2016 Mar 21;11(3):e0151109. doi: 10.1371/journal.pone.0151109.

Waumans Y, Vliegen G, Maes L, Rombouts M, **Declerck K**, Van Der Veken P, Vanden Berghe W, De Meyer GR, Schrijvers D, De Meester I. *The Dipeptidyl peptidases 4, 8, and 9 in mouse monocytes and macrophages: DPP8/9 inhibition attenuates M1 macrophage activation in mice.* Inflammation. 2016 Feb;39(1):413-24. doi: 10.1007/s10753-015-0263-5.

Declerck K, Szarc vel Szic K, Palagani A, Heyninck K, Haegeman G, Morand C, Milenkovic D, Vanden Berghe W. *Epigenetic control of cardiovascular health by nutritional polyphenols involves multiple chromatin-modifying writer-reader-eraser proteins.* Curr Top Med Chem. 2016;16(7):788-806. Review.

2015:

Szarc vel Szic, K, **Declerck, K**, Vidakovic, M, Vanden Berghe, W *From inflammaging to healthy aging by dietary lifestyle choices: is epigenetics the key to personalized nutrition?* Clin Epigenetics. 2015 Mar 25;7:33. doi: 10.1186/s13148-015-0068-2.

2014:

Milenkovic, D, Vanden Berghe, W, Boby, C, Leroux, C, **Declerck, K**, Szarc vel Szic, K, Heyninck, K Laukens, K, Bizet, M, Defrance, M, Dedeurwaerder, S, Calonne, E, Fuks, F, Haegeman, G, Haenen, G, Bast, A, Weseler, A. *Dietary flavanols modulate the*

transcription of genes assiociated with cardiovascular pathology without changes in their DNA methylation state. PLoS One. 2014 Apr 24;9(4):e95527. doi: 10.1371/journal.pone.0095527.

Szarc vel Szic K, Op de Beeck, K, Ratman, D, Wouters, A, Beck, IM, **Declerck, K**, Heyninck, K, Fransen, E, Bracke, M, De Bosscher, K, Lardon, F, Van Camp, G, and Vanden Berghe, W *Pharmacological levels of Withaferin A (Withania somnifera) trigger clinically relevant anticancer effects specific to triple negative breast cancer cells.* PLoS One. 2014 Feb 3;9(2):e87850. doi: 10.1371/journal.pone.0087850.

Conferences

Oral Presentations:

Declerck K. *Blood surrogate epigenetic biomarkers of atherosclerosis reveal common signature of inflamm-aging-disorders*. EpiChem COST meeting, Faro, Portugal, 18-20 February 2019

Declerck K. *Blood surrogate epigenetic biomarkers of atherosclerosis reveal common signature of inflamm-aging-disorders*. 10th Research Day of the FBD Faculty at University of Antwerp, Belgium, 26 October 2018

Declerck K, Suter A, Vanden Berghe W. *Systems biology characterization of immunomodulatory properties of Echinacea in monocyte cells.* 3rd Scientific Workshop POSITIVe COST: "Omics breakthroughs in the health effects of plant food bioactive", Thessaloniki, Greece, 20-21 September 2017

Poster Presentations:

Declerck K, Perez Novo C, Logie E, Chirumabmilla C, Sedigheh Takhsha F, Vanden Berghe W. Multiple myeloma therapy response to GC involves epigenetic modulation of B-cell receptor signaling. BACR Annual Meeting, Antwerp, Belgium, 1st of February 2019

Declerck K, Sedigheh Takhsha F, Vanden Berghe W. *Glucocorticoids induce DNA methylation changes in multiple myeloma cells via a GR-dependent mechanism.*

Nuclear Receptor Research Network (NRRN 2017), Leuven, Belgium, 7 November 2017.

Declerck K, Sedigheh Takhsha F, Vanden Berghe W. *Glucocorticoids induce DNA methylation changes in multiple myeloma cells via a GR-dependent mechanism*. Barcelona Conference on Epigenetics and Cancer: Epigenetic Mechanisms in Health and Disease, Barcelona, Spain, 25-26 October 2017.

Declerck K, Sylvie Remy S, Wohlfahrt-Veje C, Main MK, Schoeters G, Vanden Berghe W, Andersen HR. *Prenatal pesticide exposure interacts with a common polymorphism in the PON1 gene leading to specific changes in DNA methylation in genes associated with cardio-metabolic disease risk.*
Clinical Epigenetics Society (CLEPSO) International Meeting, Düsseldorf, Germany, 3-4 March 2016.

Declerck K, Szarc vel Szic K, Palagani A, Heyninck K, Haegeman G, Milenkovic D, Vanden Berghe W. *Epigenetic control of cardiovascular health by diet or cocoa flavanol metabolites involves multiple chromatin writer-eraser enzymes.*
EMBO conference: Chromatin and epigenetics, EMBL Heidelberg, Germany, 6-10 May 2015.