Transnational University Limburg is a unique collaboration of two universities in two countries: the University of Hasselt and Maastricht University.

UHASSELT
KNOWLEDGE IN ACTION

www.uhasselt.be
Universiteit Hasselt
Campus Hasselt:
Martelarenlaan 42 | 3500 Hasselt
Campus Diepenbeek:
Agoralaan Gebouw D | 3590 Diepenbeek

UHASSELT
KNOWLEDGE IN ACTION

Maastricht University

Faculty of Medicine and Life Sciences
School for Life Sciences

Master of Biomedical Sciences

Masterthesis

Epigenome-wide DNA methylation and mitochondrial DNA content in relation to in utero $PM_{2.5}$ exposure

Brigitte Reimann

Thesis presented in fulfillment of the requirements for the degree of Master of Biomedical Sciences, specialization Environmental Health Sciences

SUPERVISOR :
Prof. dr. Michelle PLUSQUIN

CO-SUPERVISOR :
dr. Bram JANSSEN

MENTOR :
Mevrouw Annette VRIENS

2017
2018

UHASSELT
KNOWLEDGE IN ACTION

Maastricht University

Faculty of Medicine and Life Sciences
School for Life Sciences
Master of Biomedical Sciences

Masterthesis

Epigenome-wide DNA methylation and mitochondrial DNA content in relation to in utero $PM_{2.5}$ exposure

Brigitte Reimann

Thesis presented in fulfillment of the requirements for the degree of Master of Biomedical Sciences, specialization Environmental Health Sciences

SUPERVISOR :
Prof. dr. Michelle PLUSQUIN

MENTOR :
Mevrouw Annette VRIENS

CO-SUPERVISOR :
dr. Bram JANSSEN

Table of Contents

Acknowledgements

The last eight months have been a long but fascinating journey which led from simple questions like "Where can I download R ?" to more elaborate problems like "Where can I store the huge output from the supercomputer?"

Even though the preceding years of study prepared us to some extend for the enormous undertaking that is a master thesis, I guess all of us were somewhat overwhelmed by the shear amount of complexity and consideration the planning and execution of this task required.

Without the support and guidance of many people affiliated with Hasselt University as well as family and friends I would not have come so far.

First of all, I would like to thank my supervisor professor Michelle Plusquin who always had an open ear for my ideas, helped me further when I was stuck with my analyses, and pointed me in the right direction when I was overthinking things again.

I also want to thank my co-supervisor dr. Bram Janssen for his patience in helping me understand the do's and don'ts of mediation analysis and my daily supervisor Annette Vriens for her support and help with all my questions. Further I owe special thanks to Rossella Alfano, M.D. and dr. Bianca Cox. Rossella, I want to thank you for all the time and help you provided me with, even though you had a very busy program yourself commuting between Italy and Hasselt. Everything seemed much easier when you explained it. Bianca, your help and guidance with R scripts and mediation analysis in SAS was much appreciated and often meant salvation from hours and hours of pondering. Furthermore, I also want to thank everyone from the Epidemiology group of the CMK not mentioned already for their support in many small things that added up over the last eight months.

Last but not least, I want to thank my family, my husband and children, parents and parents in law for their support at every single day for the last five years. I know, I asked a lot and gave so little in this time and will make up for it in the future.

List of Abbreviations

5hmC	5-hydroxymethylcytosine
5'UTR	5' untranslated region
8-OHdG	8-hydroxy-2’-deoxyguanosine
ACTB	Beta actin
ASDs	Autism spectrum disorders
ASK1	Apoptosis signaling kinase 1
BER	Base excision repair
BH	Benjamini Hochberg
CPT1B	Carnitine Palmitoyltransferase 1B
DE	Direct effect
DMRs	Differentially methylated regions
ENVIRONAGE	ENVIRonmental influence *ON* AGEing in early life
EWAS	Epigenome-wide association studies
FDR	False discovery rate
FWER	Family-wise error rate
HDAC	Histone deacetylase
HDAC4	Histone Deacetylase 4
HNRNPM	Heterogeneous Nuclear Ribonucleoprotein M
IARC	International Agency for Research on Cancer
IE	Indirect effect
MAD1L1	Mitotic Arrest Deficient 1 Like 1
meanbetafc	Mean ß fold change
MHC	Human major histocompatibility complex
mtDNA	Mitochondrial DNA
mtDNMT1	Mitochondrial DNA (cytosine-5)-methyltransferase 1
MTF3212	Mitochondrial forward primer from nucleotide 3212
MT-ND1	Mitochondrial encoded NADH dehydrogenase 1
mTORC1	Mammalian target of rapamycin complex 1
mTORC1	MTOR Complex 1
nDNA	Nuclear DNA
NF115	Ring Finger Protein 115
NGEMPs	Nuclear genes encoding mitochondrial proteins
no.cpgs	Number of CpGs constituting the differentially methylated region
NTCs	No template controls
PAHs	Polycyclic aromatic hydrocarbons
PCA	Principal component analysis
PM	Particulate matter
$PM_{2.5}$	Particles with an aerodynamic diameter of 2.5 μm or smaller
PROGRESS	Programming Research in Obesity, Growth, Environment and Social Stressors
PSGs	Human pregnancy-specific glycoproteins
R3319	Reverse primer from nucleotide 3319
RMTs	Arginine methyltransferases
ROS	Reactive oxygen species
RPLP0	Acidic ribosomal phosphoprotein P0
RPTOR	Regulatory Associated Protein Of MTOR Complex 1
SAM	Sadenosyl-methionine
SLE	Systemic lupus erythematosus
TET1	Tet Methylcytosine Dioxygenase 1
TSS	Transcription start site
UFPs	Ultrafine particles
UNKL	Unkempt Family Like Zinc Finger protein
vWf	Von Willebrand factor

Abstract

Background: *In utero* exposure to particulate matter (PM) is associated with adverse perinatal outcomes and health problems later in life but little is known about the underlying molecular mechanisms. Epigenome-wide DNA methylation and mitochondrial DNA (mtDNA) content of adult peripheral blood and placental tissue have already been described in literature as associated with exposure to $PM_{2.5}$. In this study we assessed whether epigenetic profiles and changes in mtDNA content in cord blood are associated with *in utero* $PM_{2.5}$ exposure and if they are related with each other.

Methods: Epigenome-wide association studies (EWAS) were performed in 187 cord-blood samples from the ENVIR*ON*AGE birth-cohort with methylation data from the EXPOsOMICS project using Infinium 450K technology. mtDNA content was determined by qPCR and maternal $PM_{2.5}$ exposure assessed by a spatial temporal interpolation method based on the CORINE land-cover data set. Associations were investigated using multiple linear mixed regression models; *p*-values were corrected for multiple testing by controlling the FDR. Furthermore, an integration with gene expression data and pathway analyses was performed for the 20 top hits according to *p*-values for the association with $PM_{2.5}$ and mtDNA content and the overlaps in the results were investigated. On the level of differentially methylated regions (DMRs), associations were examined using the DMRcate and bumphunter packages in R. Finally, a mediation analysis was performed for the effects of $PM_{2.5}$ on CpG site specific methylation through changes in mtDNA content.

Results: mtDNA content was negatively associated with *in utero* $PM_{2.5}$ exposure in a model adjusted for gestational age, newborn's sex and smoking during pregnancy and number of platelets (β = -0.038 ± 0.0076, $p < 0.0001$). Six CpG sites, five of which were annotated showed a significant association (q-value < 0.2) with $PM_{2.5}$ in an EWAS, all being hypermethylated. The most significant CpG site was located on a CpG island in an intron of *VWA1*, encoding the 'von Willebrand factor' (vWf). This CpG site was also present among the 70 DMRs identified by DMRcate for a significant (z-value <0.2) association with $PM_{2.5}$. The most significant DMR (Stouffer z-value = 0.01) was located on an exon of *RPTOR,* which encodes a component of the mTORC1 complex that prevents autophagy. The corresponding CpG sites were part of a CpG island and showed hypermethylation. With the bumphunter algorithm an overlap of four DMRs, located on *STIL, HLA-DRB6, CCKBR* and *CPT1B* was found within the top 20 of both associations. Pathway analysis revealed an overlap of 13 enriched pathways involved in histone modifications, mitosis, cell cycle and systemic lupus erythematosus. For the effect of prenatal $PM_{2.5}$ exposure on the methylation status of a CpG site located on *HNRNPM*, a significant mediation by mtDNA content was found.

Conclusions: With the results of this study we provided evidence not only for significant associations between prenatal $PM_{2.5}$ exposure and mtDNA content as well as epigenome wide DNA methylation but for the first time also demonstrated a link between these two biomarkers of early biological effect. This could present a step towards a better understanding of the underlying molecular mechanisms that link prenatal $PM_{2.5}$ exposure with adverse pregnancy outcomes and later life health problems. The identification of common genes and pathways involved in histone modifications, mitosis, cell cycle and innate immunity presents potential common mechanisms between differential DNA methylation and mtDNA content in relation with prenatal $PM_{2.5}$ exposure and offers possible directions for future research.

Epigenome-wide DNA methylation and mitochondrial DNA content in relation to *in utero* $PM_{2.5}$ exposure

1 Introduction

1.1 Adverse effects of particulate matter on human health

Ambient air pollution originating from combustion sources is composed of different gases and particulate matter (PM), a heterogenous mixture of airborne particles with an aerodynamic diameter of 10 µm or smaller (1). Constituents of PM include oxidizing compounds, such as transition metals, and organic substances, such as polycyclic aromatic hydrocarbons (PAHs) (1, 2). Due to their limited size, particles with an aerodynamic diameter of 2.5 µm or smaller ($PM_{2.5}$) are inhaled deeply into the lungs where they become lodged in the alveoli and provoke inflammatory reactions. The resulting cytokines and chemokines can affect other organ-systems and cause cellular damage there (3). The fraction of ultrafine particles (UFPs) with an aerodynamic diameter of 0.1 µm or smaller can even penetrate the lung epithelium, reach the blood stream and distribute systemically (4). Once inside the body the particles can spread to many organs and tissues and interact directly with cells, provoking adverse effects.

Exposure to ambient air pollution represents a major risk to human health, with an estimated three million deaths per year, predominantly occurring in the low- and middle-income countries of South-East Asia and the Western Pacific regions (5). The World Health Organization recommends air quality guidelines of $PM_{2.5}$ concentrations of less than 10 µg/m^3 annual mean, and 25 µg/m^3 24-hour mean concentrations. In 2014 about 90% of the population living in cities was exposed to $PM_{2.5}$ concentrations exceeding these guidelines (5). This is problematic considering that even at very low concentrations acute (6, 7), as well as long-term (8, 9) exposure to $PM_{2.5}$ is quantitatively related to increased mortality or morbidity. Cardiovascular and respiratory diseases are among the conditions most often associated with $PM_{2.5}$ exposure (8, 10). Additionally, the exposure to $PM_{2.5}$ has been linked with lung cancer incidence and mortality in European as well as US-American cohorts (11, 12) which prompted the International Agency for Research on Cancer (IARC) in 2013 to classify outdoor air pollution as carcinogenic to humans (Group 1) (13). The adverse effects of air pollution are already measurable in newborns as a consequence of intrauterine exposure. Prevailing epidemiological evidence suggests, that maternal $PM_{2.5}$ exposure is correlated with adverse perinatal outcomes like preterm birth (14-16) and low birth weight (16, 17).

According to the 'Developmental Origins of Health and Disease' hypothesis prenatal environmental exposures can also increase disease susceptibility later in life and may therefore result in early childhood diseases as well as in adverse health conditions in middle age (18). Previous research showed indeed an association between *in utero* $PM_{2.5}$ exposure and early childhood cancer (19, 20). Other studies were able to demonstrate a relationship with autism spectrum disorders (ASDs) or suggested a negative impact on the neuropsychological development of children (21). Furthermore, respiratory ailments in children such us a decreased lung function (22, 23) and an increased susceptibility to respiratory infections (24) were linked to *in utero* $PM_{2.5}$ exposure. Evidence from an experimental study on mice established a relationship between *in utero* exposure to $PM_{2.5}$ and cardiac failure in adult mice which corroborates the hypothesis that the influence of the *in utero* exposure can even persist during adulthood and impact the health status later in life (25).

1.2 Modes of action

1.2.1 Transplacental transfer

One possible way $PM_{2.5}$ may exerts its influence on the fetus, is by passage through the placental barrier. Particles with a diameter up to 240 nm were able to cross this barrier in a perfusion model of the human placenta (26). Additionally, soluble organic chemicals like PAHs contained in $PM_{2.5}$ showed significant correlations between maternal and umbilical cord serum concentrations, suggesting transplacental transfer from mother to fetus (27, 28). Likewise, the fetus could be affected by the pro-inflammatory response of the mother; placental transfer from the maternal to the fetal side has been reported for interleukin 6 in a perfusion model (29). Once passed the placental boundary, the components of $PM_{2.5}$ and pro-inflammatory cytokines may cause genetic and epigenetic changes in the fetus.

1.2.2 Effects on the genetic level: mitochondrial DNA damage and content

Especially on the genetic level oxidative stress appears to be a prominent mechanism of $PM_{2.5}$ toxicity and mitochondrial DNA (mtDNA) one of its main targets. mtDNA consists of a closed circular molecule that encodes for the 13 components of the oxidative phosphorylation complexes in humans. Its proximity to the electron transport chain at the inner mitochondrial membrane, may result in an increased risk of DNA damage. Not only bears the inner mitochondrial membrane the origin of most cellular reactive oxygen species (ROS) but it also accumulates lipophilic, positively charged molecules from exogenous sources (30). In contrast to nuclear DNA (nDNA), mtDNA lacks protective features like histones and efficient repair mechanisms (30). Furthermore, damage and mutations in the mtDNA are much more prone to cause functional consequences than damage in nDNA because almost all base pairs of the mtDNA are coding except the sequence at the D-loop which contains the promoters for the two complementary DNA strands and the origin of replication (31). The significant impact of mtDNA damage and mutations is underlined by the fact that mitochondrial dysfunction is linked with several age-related and neurodegenerative diseases of considerable public health relevance, such as cancer (32), diabetes (33), Alzheimer's (34), Parkinson's (35), and Huntington's (36) disease which makes mtDNA damage also suitable as an effect biomarker (37). The higher susceptibility of mtDNA to damage and mutations compared to nDNA can also be explained by the difference in repair mechanisms. In contrast to small lesions such as 8-OHdG, which are efficiently repaired by base excision repair, larger lesions like adducts often remain unrepaired in mtDNA due to the lack of adequate nucleotide excision repair (30). In *in vitro* experiments with human fibroblasts, mtDNA has therefore been shown to accumulate more polymerase-blocking lesions of the sugar–phosphate backbone (e.g., strand breaks, abasic sites and DNA adducts) as compared with nDNA (38).

There are different ways in which $PM_{2.5}$ can cause oxidative stress to mtDNA. The incorporated particles can generate reactive oxygen species (ROS) directly on their surface, or after releasing metal ions and soluble organic chemicals (39). The oxidation of guanine followed by the formation of 8-hydroxy-2′-deoxyguanosine (8-OHdG) is the most frequent form of oxidative DNA damage in nuclear as well as in mitochondrial DNA (40). Other DNA lesions regularly caused by oxidative stress are strand breaks, base lesions and cross-links (41). The initial DNA damage caused by $PM_{2.5}$ is amplified by the altered function of mitochondria and the activation of inflammatory cells capable of generating even more ROS (39, 42). Generally, the reactivity and therefore potential to cause oxidative stress

seem to be inversely related to the size of the PM particles due to the increase in surface area to volume ratio (2). $PM_{2.5}$ which comprises the fine fraction of particulate matter therefore has a higher potential for causing oxidative stress compared to the coarse fraction of particulate matter after systemic distribution by the bloodstream. Additionally, non-oxidative DNA damage is caused by PAHs adsorbed onto respirable air particles. After inhalation of the $PM_{2.5}$ particles the PAHs can form reactive epoxides that covalently bind to DNA, creating PAH-DNA adducts (2, 43). PAH-DNA adducts increase the probability of gene mutations and are associated with various forms of cancer which makes them an established markers of cancer risk (1, 44). In a previous study an association between prenatal $PM_{2.5}$ exposure and mtDNA damage in cord blood (45) could not be demonstrated. This study determined mitochondrial 8-OHdG by quantitative polymerase chain reaction (qPCR) and its association with $PM_{2.5}$ was studied for various time windows during pregnancy, showing an inverse trend for the last trimester. Because of the comparably efficient repair of small DNA lesions in mtDNA the use of 8-OHdG as an marker of mtDNA damage could still result in an underestimation. 8-OHdG has even been shown to be less prevalent in liver mtDNA of aged animals as compared with nDNA (46) demonstrating the limits of its use as a marker of mtDNA damage. An alternative marker which allows the evaluation of longer periods of exposure and long lasting functional consequences is the determination of mtDNA content. The accumulation of mtDNA damage and mutation caused by oxidative compounds results in an increasing dysfunction of mitochondria. In a first instance the lost energetical capacity is compensated by an increase of the mtDNA copy number but in case the accumulated damage exceeds the mitochondrion’s compensatory mechanisms a decrease in mtDNA content occurs (47, 48). Accordingly, low mtDNA content constitutes an established marker of mitochondrial damage and dysfunction (49). The assessment of mtDNA content in cord blood can therefore serve as a proxy for mtDNA damage sustained during pregnancy. Previous research has already investigated the association between prenatal $PM_{2.5}$ exposure and mtDNA content in placental tissue (50) and found evidence for a negative relationship for the entire pregnancy which was most pronounced for the last trimester. In another study this association was confirmed in cord blood for the last weeks of the third trimester when fitting distributive lag models (51) however the mother-newborn pairs were recruited in the Mexican 'Programming Research in Obesity, Growth, Environment and Social Stressors' (PROGRESS) study comprising a majority of low-income workers with less than 12 years of schooling (77%) and a low socioeconomic status. It remains to be seen, if the above mentioned findings can be replicated for other socio-economic backgrounds and exposure ranges.

1.2.3 Effects on the epigenetic level: epigenome-wide DNA methylation

Epigenetic changes such as epigenome wide DNA methylation could also mediate some of the effects of *in utero* $PM_{2.5}$ exposure. The most sensitive period for the fetal programming is directly after the fertilization when the epigenome of the early embryo is established (52). DNA methylation at the carbon-5 position of cytosine in CpG dinucleotides is the most investigated epigenetic event and plays a crucial role in responding to the environment through changes in gene expression (53). As methylation occurs on CpG islands, and guanine is the nucleotide base that is most sensitive to oxidation, it is apparent that ROS have the potential to cause alterations to the DNA methylation pattern. For mtDNA methylation it has been demonstrated that oxidative stress can modulate mitochondrial DNA (cytosine-5)-methyltransferase 1 (mtDNMT1) gene expression via p53 and hereby influences the mitochondrial DNA methylation and subsequently mitochondrial gene transcription (54).

Furthermore, ROS can cause various oxidative DNA lesion which influence the amount of methylation. 5-hydroxymethylcytosine (5hmC) is a physiological product of oxidation, and only poorly recognized by DNA methyltransferases resulting in DNA demethylation (55). Oxidation of guanine to 8-oxoguanine, or of 5-methylcytosine to 5hmC, significantly inhibits the binding capacity of the methyl CpG-binding domain-proteins with the corresponding CpGs (56). Additionally, metals such as cadmium which are contained in $PM_{2.5}$ can interact with the DNA binding domain of the methyltransferases or deplete the cell of the methyl-donor sadenosyl-methionine (SAM) causing global DNA hypomethylation (57). Methylated CpG (mCpG) sequences are also preferred targets for PAHs which form guanine adducts and induce G to T transversion mutations (57). Recent studies, investigating the link between $PM_{2.5}$ exposure and the epigenome-wide methylation, found associations with individual differentially methylated CpG sites in adults (58, 59). In the first of these two studies significant relationships between epigenome-wide methylation at individual CpG sites and one-year moving averages of $PM_{2.5}$ species were found in a longitudinal cohort study, in the latter a significant association between individual CpG sites and $PM_{2.5}$ exposure for the independent populations of two cohorts in a meta-analysis were demonstrated. Previous studies on *in utero* exposure to air pollution have concentrated on the effects on global (60) or mitochondrial methylation (50) in placental tissue. Examining the effect of $PM_{2.5}$ on global methylation, a significant inverse association for the entire pregnancy, driven by the values of the first trimester, especially during the implementation phase, were demonstrated. Likewise, for the methylation of two mtDNA loci, a positive association was shown for the entire pregnancy as well as for the first trimester.

Possible effects of *in utero* $PM_{2.5}$ exposure on epigenome-wide methylation in cord blood have, to our knowledge, not been investigated yet. An epigenome wide analysis offers, compared to the investigation of global and region-restricted methylation the advantage of an unbiased approach, able to identify individual CpGs with an altered methylation status as well as differentially methylated regions (DMRs). DMRs are genomic regions which show different levels of DNA methylation under distinct biological conditions. They combine spatial information from multiple nearby methylation sites that often fulfill a common functional purpose. The analysis of individual CpGs with an altered methylation status or DMRs can in turn be linked to changes in gene expression which provide insights into biological processes and disease pathways (61).

1.2.4 Possible interactions between the genetic and epigenetic level

The scope of the above-mentioned studies was limited to the investigation of either the associations between $PM_{2.5}$ exposure, and mtDNA content or $PM_{2.5}$ exposure and epigenome-wide DNA methylation. The relationship between the two proposed molecular mechanisms of $PM_{2.5}$ toxicity itself has not yet been investigated and the influence can be bi-directional. The genetic and epigenetic level are often linked, for instance, when DNA lesions such as 5hmC interfere with the ability of methyltransferases to interact with DNA, hence causing hypomethylation of cytosine residues at CpG sites. An important aspect which could be of utmost importance in the interplay between the genetic and epigenetic level is the effect of external influences on the epigenetical landscape via the energetical status of the cell. Mitochondria as the 'powerhouses' of the cell anticipate to the energy demands of the processes required for DNA methylation and histone modification. Changes in the function, efficiency and biogenesis of mitochondria caused by mtDNA damage and consequent decrease in mtDNA content could therefore affect the epigenetic level. Likewise, the differential

methylation of promoter regions could alter the expression of general repair- and antioxidative genes involved in DNA-damage-control and of more specific nuclear genes encoding mitochondrial proteins (NGEMPs). Since knowledge about the underlying mechanisms of $PM_{2.5}$ toxicity is limited, one goal of this study was to investigate possible associations between the two biomarkers epigenome-wide methylation and mtDNA content in cord blood and the pathways that link them.

1.3 Objectives

For the purpose of this study, we used cord blood samples from the ongoing, population-based, prospective ENVIR*ON*AGE (ENVIRonmental influence *ON* AGEing in early life) birth cohort to investigate how epigenome-wide methylation and mtDNA content in newborns are affected by maternal exposure to $PM_{2.5}$ during pregnancy. We hypothesized that we can identify CpGs and differentially methylated regions (DMRs) associated with prenatal $PM_{2.5}$ exposure and mtDNA content. For this purpose, we considered different molecular outcomes as depicted in Fig. 1. To determine if prenatal $PM_{2.5}$ exposure affects mtDNA content we first applied a multiple linear regression model adjusting for a limited scope of known potential confounders (model 1). In a second step we compared our findings with the outcome of a second model with additional variables (model 2). To determine sensitive exposure windows, we also fitted an exposure–lag–response model in which the three trimester-exposures were present as independent variables in the same regression model. Furthermore, we employed epigenome-wide association studies (EWAS) to identify CpG sites with differential methylation associated with (i) prenatal $PM_{2.5}$ exposure and (ii) mtDNA content. The purpose of this analyses was to investigate if prenatal $PM_{2.5}$ exposure induces changes in the methylation status of individual CpG sites and to further examine if these potential changes could in turn be linked to changes in mtDNA content. The EWAS were conducted with model 1 in a multiple linear mixed regression model to identify significant associations and subsequently tested in model 2 for comparison. For the 20 top hits according to *p*-values for the association with $PM_{2.5}$ we also fitted an exposure–lag–response model as described above to determine the most vulnerable periods of the pregnancy. To obtain insights into the biological relevant consequences of the associations an integration with gene-expression data and consequent pathway analysis was performed for the 20 top hits according to *p*-values for both EWASs and the outcome was examined for overlap. To validate our findings, we also studied the associations between $PM_{2.5}$ and mtDNA content with epigenome-wide methylation on the level of DMRs. For this purpose, we used two different R packages to obtain DMRs and compare the outcomes for overlap. Finally, we also performed mediation analyses to investigate in how far changes in mtDNA content may mediate the association between prenatal $PM_{2.5}$ exposure and differential methylation at certain CpG sites.

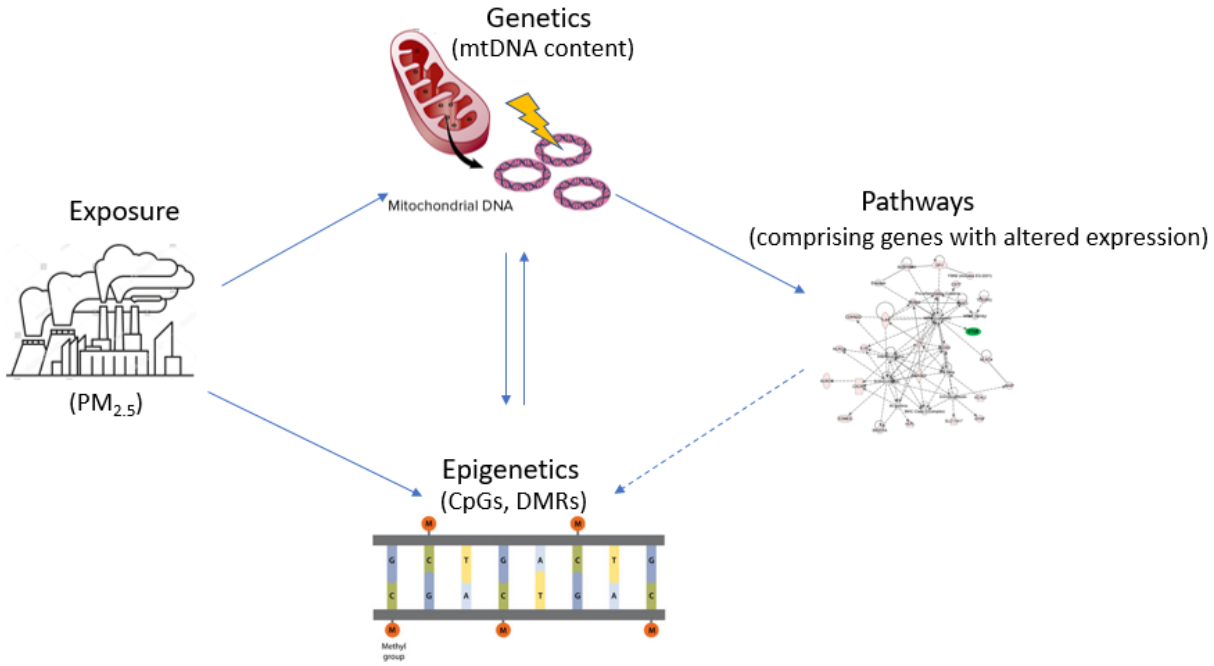


Figure 1. Illustration of the structure and relationships between the research objectives. $PM_{2.5}$, particulate matter with a diameter ≤ 2.5 μm; DMRs, differentially methylated regions; mtDNA, mitochondrial DNA.

2 Methods

2.1 Study population

Initially this study included 200 mother-newborn pairs with only singleton newborns originating from the ENVIR*ON*AGE birth cohort situated in the province of Limburg, Belgium (62). The mothers were recruited between 2014 and 2015 at arrival in the delivery ward of the East-Limburg Hospital in Genk. Each of the mother-newborn pairs provided maternal and umbilical cord blood. After delivery, the mothers filled out a questionnaire regarding information about in-house environment, education, occupation, health status, smoking, and life-style habits. Perinatal parameters such as newborn's sex and birth date were collected by the hospital personnel. Data about the methylation status of the CpG sites was retrieved in the framework of the EXPOsOMICS project (FP7), which aims to access the totality of environmental exposures from conception onwards and links them to biochemical and molecular changes in the body (63).

Due to missing information on $PM_{2.5}$ exposure three mother-newborn pairs had to be excluded from the analysis of epigenome-wide DNA methylation and $PM_{2.5}$ exposure. For another 10 mother-newborn pairs data on education of the mother were missing. Therefore, 187 mother newborn pairs with complete covariate data were included in this analysis. For the study of the association between mtDNA content and $PM_{2.5}$ as well as mtDNA content and DNA methylation, mtDNA content measurements for 192 of the 200 mother newborn pairs were available. Again 13 pairs had to be excluded due to absent data of $PM_{2.5}$ exposure and education of the mother and additionally three pairs because of missing measurements of platelet count. This analysis was consequently performed with 176 samples.

The present study was conducted according to the principles outlined in the Helsinki Declaration (64) and approved by the Ethical Committee of Hasselt University and the East-Limburg Hospital in Genk, Belgium. Written informed consent was obtained on forehand from all participating mothers.

2.2 Air pollution exposure assessment

Prenatal $PM_{2.5}$ exposure ($\mu g/m^3$) was determined based on the residential address of the mother. Therefore, a spatial temporal interpolation method was employed that combines pollution data collected in the official governmental stationary monitoring network and land-cover data retrieved from satellite images (CORINE land-cover data set) with a dispersion model to estimate the exposure (65). This interpolation method provides a high-resolution grid of an average size of 25 × 25 m, explaining more than 80% of the temporal and spatial variability in the daily $PM_{2.5}$ measurements. Besides calculating exposure during the entire pregnancy as the mean of all pregnancy days, averages were determined for the three pregnancy trimesters to allow for the identification of critical exposure windows. Accordingly, the first trimester (week 1–13), second trimester (week 14–26) and third trimester (week 27– delivery) were predefined by estimating the date of conception based on ultrasound data. If address-changes occurred in the course of the pregnancy these were taken into account when calculating the exposure.

2.3 Mitochondrial DNA content

DNA was extracted from white blood cells of the buffy coat after centrifugation of the samples using the QIAamp DNA mini kit (Qiagen, Inc., Venlo, the Netherlands) according to the manufacturer's instructions. The quantity and purity ratio (A260/280 and A260/230) of the extracted DNA was assessed by spectrometric analysis using the Nanodrop 1000 spectrophotometer (Isogen, Life Science, Belgium) and the DNA was diluted to a final concentration of 5 ng/μL in RNase free water. Until further processing the extracted DNA was stored at -80°C. mtDNA content was calculated by calculating the ratio of two mitochondrial gene copy numbers [mitochondrial forward primer from nucleotide 3212/ reverse primer from nucleotide 3319 (*MTF3212/ R3319*) and mitochondrial encoded NADH dehydrogenase 1 (*MT-ND1*)] to two single-copy nuclear control genes [acidic ribosomal phosphoprotein P0 (*RPLP0*) and beta actin (*ACTB*)] using a quantitative real-time polymerase chain reaction (qPCR) assay. Briefly, 7.5 μL master mix, consisting of 5 μL/reaction Fast SYBR® Green I dye 2× (Applied Biosystems), forward and reverse primer (each 0.3 μL/ reaction), and RNase free water (1.9 μL/reaction) were aliquoted into the wells of a MicroAmp® Fast Optical 96-Well Reaction Plate. To each well 2.5 μL from one of the diluted DNA samples was added to obtain a final volume of 10 μL per reaction. To account for inter-run variability and possible DNA contamination six inter-run calibrators (IRCs) and two no template controls (NTCs) were run together with the samples on each reaction plate. The thermal cycling conditions were as follows: (i) for the activation of the AmpliTaq Gold® DNA-polymerase 20 sec at 95°C, (ii) for denaturation 40 cycles of 1 sec at 95°C and (iii) for annealing and extension of the PCR products 20 sec at 60°C. Melting curve analyses were used at the end of each run to confirm the specificity of the reaction and absence of primer-dimers. Calculations of the cycle threshold (CT) values for the two mitochondrial genes were performed using the 'qBase' software (Biogazelle, Zwijnaarde, Belgium) which performs a normalization step relative to the nuclear reference genes by applying the ΔΔCT method and also taking the inter-run calibrators into account.

2.4 Epigenome-wide methylation

Cord blood samples were collected in BD Vacutainer® Plus Plastic K2EDTA Tubes (BD, Franklin Lakes, NJ, USA) immediately after delivery. The collected cord blood was centrifuged (3,200 rpm for 15 min) to retrieve buffy coat which was instantly frozen at –80°C. After thawing and extraction with the QIAamp DNA mini Kit (Qiagen Ltd, Manchester, UK) DNA was bisulphite-converted using the Zymo EZ DNA methylation™ kit (Zymo, Irvine, CA, USA). To determine the epigenome-wide DNA methylation profile of the cord blood samples the DNA was then hybridized to Illumina Infinium Human Methylation450K BeadChip arrays (66) and scanned using the Illumina HiScanSQ system. After background subtraction using Ilumina GenomeStudio raw intensity data were submitted to pre-processing, including normalization, using in-house software within the R statistical computing environment. Furthermore, quality control of samples was carried out and failed samples were excluded on the basis of Illumina's detection *p*-value of $p > 0.001$ and low bead count (<3 beads). For probes using the Infinium II design additional background subtraction and dye bias correction were performed. Methylation levels at each CpG locus were expressed as Beta-values, the ratio of signal intensity originating from methylated CpGs over the sum of methylated and unmethylated CpGs. Finally, data was trimmed for outliers with values larger than 3 interquartile ranges below the first quartile or above the fourth quartile.

2.5 Statistics

2.5.1 Epigenome-wide methylation studies

The methylations status of individual CpG sites in relation to *in utero* $PM_{2.5}$ was modelled as dependent variable in a multiple linear mixed regression model with the Beta-value representing the percentage of methylation difference for every unit ($\mu g/m^3$) increase of air pollutant. Batch effects caused by the different chips and chip-positions were accounted for by treating them as random effects variables while other possible confounders were considered as fixed effects variables. An evaluation of critical confounders to be added to an initial basic model was performed only for the association between epigenome-wide methylation and *in utero* $PM_{2.5}$ exposure. The resulting covariates were then applied also for associations with mtDNA content to ensure comparability of the outcomes in later analyses. Therefore, in a first step, a basic model containing the two random effects variables chip and position on the chip and the fixed effects variables newborn's sex and maternal smoking status was used to evaluate the impact of possible confounders and subsequently build a relevant first model (model 1). The basic model also took into account an adjustment for blood cell composition which was achieved by using the de-convolution approach proposed by Bakulski et al. (67).

In order to perform the evaluation of critical confounders, principal component analysis (PCA) was applied and the association between the first 10 principal components and the different variables was assessed in multiple linear mixed regression models, correcting for chip and position as random effects variables with the principal components as depended variable and the variable in question as experimental variable. The resulting *ß*-coefficients and *p*-values were then evaluated by means of a heatmap. In addition, QQ plots were analyzed after performing linear mixed model regression adding the variable in question as experimental variable to the basic model. Furthermore, after running the linear mixed regression model adjusting for the variables of the basic model in the complete set of CpGs the 20 top hits according to *p*-values for the association with $PM_{2.5}$ were selected and used to analyze the influence of potential confounders. Therefore, these variables were added each at a time as additional covariate to the basic model and the resulting *p*-values were plotted for visual evaluation. Model 1 additionally included the variable gestational age based on ultrasound examinations. To correct for multiple testing the method of Benjamini Hochberg (BH) was applied and individual CpGs sites were considered significant with a *q*-value < 0.2. In a following step the top 20 CpGs with the lowest *p*-value were selected for further investigation of the association with *in utero* $PM_{2.5}$ exposure in a second model (model 2) with additional covariates including maternal age ethnicity, parity, season of conception, maternal education, birthweight maternal pre-pregnancy BMI and apparent temperature in the last week of pregnancy.

To explore potentially critical exposures windows during gestation, $PM_{2.5}$ concentrations were calculated for each of the three trimesters of pregnancy. The associations of the 20 CpG sites most significantly associated with $PM_{2.5}$ in the EWAS were assessed in an exposure–lag–response model with the covariates for model 1 and model 2. All three trimester exposures fitted as independent variables in the same regression model.

An EWAS was also applied for the association with mtDNA content in order to examine possible relationships between $PM_{2.5}$ and mtDNA induced epigenome wide methylation changes in subsequent steps of the study. Therefore, the residuals of the $\log_{10}$ transformed mtDNA content and the platelet count were calculated and used as the independent variable in a multiple linear mixed regression model with the same covariates employed as in the EWAS for $PM_{2.5.}$ Here also the 20 top hits according to *p*-values for the

association with mtDNA content were selected to further compare the *ß*-coefficients and *p*-values between model 1 and model 2.

2.5.2 Gene expression and pathway analyses

For the top 20 CpG sites most significantly associated with either $PM_{2.5}$ or mtDNA content in the EWAS' the correlation with the full set of transcripts (n=29,164) from gene-expression data (Agilent 4 x 44K Whole Human Genome microarray, design ID 014850) from the same set of subjects was determined. For the 187 samples used in the analysis with $PM_{2.5}$ 179 samples with corresponding transcripts were available and for the 176 samples used in the analysis with mtDNA content 168 samples with corresponding transcripts were present. For the purpose of this investigation, the corresponding Beta-values of the CpGs sites in question were first corrected for the random batch effects of different chip and chip-positions by using a linear approach proposed previously (68) to retrieve the residuals. The observed gene expression values of the transcripts were treated the same way to correct for random batch effects of different hybridization dates. Those CpG–transcript pairs with a Bonferroni 5% significance level below 0.05 / (20 × 29,164) = 8.6e-8, were considered significant. The corresponding EntrezGeneID number of the genes allocated to the significant transcripts were uploaded into the online pathway analysis tool, DAVID 6.8 (https://david.ncifcrf.gov) (69), for the identification of enriched pathways against the default background of all genes expressed in Homo sapiens.

2.5.3 mtDNA content and $PM_{2.5}$

To investigate the association between *in utero* $PM_{2.5}$ exposure and mtDNA content, a multiple linear regression model was applied. Because all samples were processed in one batch no adjustment for batch effects was performed. The values for mtDNA content were $\log_{10}$ transformed to ensure normality of the data. The analysis was performed for model 1 and model 2 only using number of platelets instead of white blood cell composition as covariate. Possible collinearity of variables was examined by calculating the variance inflation factor. A formal test for normality of the data from independent variables was performed according to Shapiro-Wilk and additionally by evaluation of histograms. In case data was lacking normality, a sensitivity test was performed. To query the most sensitive time windows during gestation, all three trimester exposures were fitted as independent variables in a exposure–lag–response model, consistent with the analysis of DNA methylation.

2.5.4 Differentially methylated regions

Multiple linear regression was used to assess the associations between *in utero* $PM_{2.5}$ exposure as well as mtDNA content and DMRs in cord blood. Since none of the various method for the identification of DMRs has been widely accepted as the gold-standard, two methods, DMRcate and bumphunter (70, 71) which work under different statistical assumptions were selected for this study. The analysis of associations between in utero $PM_{2.5}$ exposure and DMRs as well as mtDNA content and DMRs was performed using both procedures and DMRs were considered significant if they reached a Stouffer z-value < 0.2 for DMRcate, or a family-wise error rate (FWER) < 0.2 for bumphunter. To ensure comparability between the outcomes the first steps of the pre-processing, before the actual implementation of the specifying algorithm, were applied in parallel for both methods. Missing methylation Beta-values were imputed with values of the k nearest neighbors using an Euclidean metric provided by the Bioconductor package 'impute'.

Consequently, the retrieved values were logit transformed to the M-value scale for better compliance with the modeling assumptions (72). Cross-reactive probes already identified by Chen et al. [2013] (73) and probes mapped to the sex chromosomes were removed using the 'RmSNPandCH' function of the DMRcate package. Additionally, probes within two nucleotides or closer to a SNP that have a minor allele frequency greater than 0.05 were filtered out because they often show a different distribution to those at greater distance (73). Batch effects due to the use of different chips were corrected using the 'correctBatchEffect' function of the package BEclear (74), which performs the whole process of searching for batch effects in a matrix of beta values and automatically correcting them, based on latent factor models. Subsequently, a design matrix was built, taking into account an *a priori* selected panel of covariates conform with model 1.

For DMRcate the matrix of M-values was annotated first with information about the genomic position of the probes using the ilmn12.hg19 annotation (75) and UCSC Genome Browser. In the process of finding DMRs itself DMRcate is agnostic to all annotations except for spatial ones like chromosomal coordinates. Then, like in the analysis of individual CpG sites, a limma linear model with empirical Bayes adjustment is fitted for each individual CpG site. Subsequently the individual CpG sites are combined by the 'dmrcate' function and modeled by Gaussian Kernel smoothing. Notably for this procedure, DMRcate uses unsigned weights (limma's t^2s) to pass to the kernel estimator were the estimates are calculated according to the formula: $K_{ij} = exp\left(\frac{-[x_i - x_j]^2}{2\sigma^2}\right)$, with the Gaussian kernel weights represented by K_{ij} for the F statistics, Y_i at the locations x_i, and the kernel scale factor σ proportional to the bandwidth λ. By this, two estimates, one weighted and one not, are derived. For each chromosome these two smoothened estimates are subsequently compared via a Satterthwaite approximation and a significance test is conducted. Probes significant after FDR correction and within a distance of λ nucleotides to each other are then grouped into a region. For this study the default smoothing parameters with bandwidth λ = 1, 000 bp and scaling factor C = 2 (kernel size = 500 bp) were applied and a minimum FDR q-value < 0.2 for the CpG sites constituting a DMR was assigned.

Bumphunter also uses smoothed methylation values to discover DMRs but in contrast to DMRcate the bumphunter algorithm first defines clusters of probes and tests for significance afterwards. Another difference with DMRcate is that bumphunter works under the assumption that methylation changes of CpGs within a given region must always be in the same direction. Consequently, signed weights are passed on to the smoothing process, which can cause a loss of biologically significant results when positive and negative signs cancel each other out. Bumphunter first computes a t-statistic for each genomic location by applying a regression model for the logit-transformed methylation measurements against the variable of interest. The estimated slope $\hat{ß}_{(ij)}$ is retained and loess smoothing is applied to create clusters. A designated candidate region is subsequently formed by clusters of nearby probes for which all the t-statistics exceeded a predefined cut-off threshold chosen by the program with a default of 0.99. Finally, permutation or bootstrapping procedures that construct null distributions for each candidate region are employed. Because in scenarios with more than one covariate the use of the permutation approach is not recommended (71) 'bootstrapping' (B=1000) was performed in our study to assess uncertainty and create null distributions for the candidate regions. To be considered as a significant DMR, candidate regions in this study had to contain at least two differentially methylated CpG sites and FWER < 0.2.

2.5.5 Mediation analysis

To investigate if the association between prenatal $PM_{2.5}$ exposure and DNA methylation is mediated by mtDNA content we performed a mediation analysis. Mediation analysis examines the mechanisms that control the observed relationship between exposure and outcome and explores their association with a third variable, the mediator. The hypothesis in this approach is therefore that besides a tentative direct change in the outcome ($ß_1$) the exposure variable also causes a change ($ß_2$) in the mediator, which successively causes a change ($ß_3$) in the outcome (Fig. 2). The role of the mediator is to give insights into the relationship between exposure and outcome (76). As an extension of the initial approach proposed by Baron and Kenny (1986) (77) the counterfactual framework works with the concept of decomposition of a total effect into direct and indirect effects (78) which allows the specification of the weights of different variables. The direct effect (DE) describes the exposure effect (X) on the outcome (Y) while holding the mediator variable (M) fixed. The indirect effect (IE) covers the effect of changes in the exposure on the outcome which act through changes in the mediator. DE and IE can therefore be described as:

$$DE = E\left(Y_{x,\ M_{x^*}}\right) - E\left(Y_{x^*,\ M_{x^*}}\right) = E\left(Y_{x,\ M_{x^*}}\right) - E\left(Y_{x^*}\right)$$

$$IE = E\left(Y_{x^*,\ M_x}\right) - E\left(Y_{x^*}\right) = E\left(Y_{x^*,\ M_x}\right) - E\left(Y_{x^*,\ M_{x^*}}\right)$$

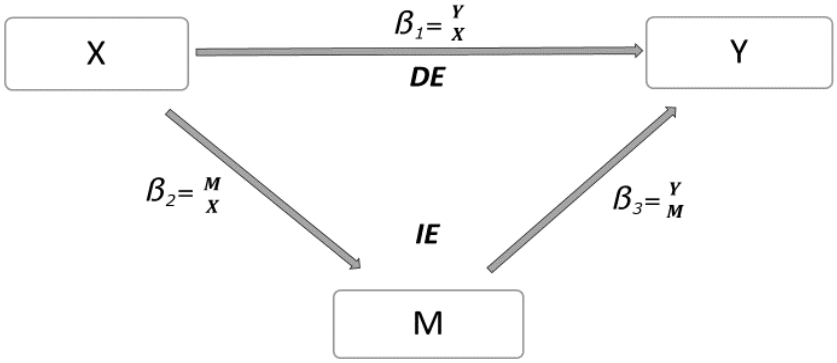


Figure 2. Relationship between exposure (X), outcome (Y), and mediator (M) in a mediation analysis as proposed by Baron and Kenny (1986), depicted as a directed acyclic graph (DAG). The direct effect (DE) and indirect effect (IE) according to the counterfactual approach are also shown.

In order to qualify as a mediator a variable has to show significant association with the exposure variable and with the outcome variable. The association between the exposure and the outcome variable must also be significant for M to be a mediator except for inconsistent mediation models, where at least one mediated effect has a different sign than the other effects (79). The estimation of the DE and IE demands two additional requirements. For the assumption to hold true, no unmeasured confounders influencing the association triangle between exposure, outcome and mediator are allowed (78). Before performing the mediation analysis, we therefore tested in a multiple linear mixed regression model only correcting for batch effects, if CpG sites identified in the EWAS as significantly associated with $PM_{2.5}$ exposure were also related to mtDNA content to fulfill the assumptions for mediation analysis. The resulting CpG sites with a *p*-value < 0.05 were then subjected to a mediation analysis for model 1 and model 2.

In the interpretation of the results of the analysis we were faced with the phenomena of inconsistent mediation (80). Because the proportion of mediation is described by (IE/DE+IE) we followed the suggestion of Alwin & Hauser (1975) and took the absolute values of the DE and IE to calculate the proportion (81). Statistical analyses were conducted using the SAS statistical package, version 9.4 (SAS Institute, Cary, NC, USA) together with SAS macros provided by Valeri et al. (79)

3 Results

3.1 Population characteristics

Demographic characteristics and perinatal factors of (i) the 187 mother-newborn pairs included in the analysis of epigenome-wide methylation and prenatal $PM_{2.5}$ exposure and (ii) the 176 mother-newborn pairs included in the analysis of epigenome-wide methylation and mtDNA content are reported in Table 1.

In this study around 90% of the newborns were Europeans of Caucasian ethnicity among them 48% girls. The mean (±SD) gestational age was 39.1 (±1.6/1.7) weeks and the mean birthweight was around 3.4kg (±476/462 Mean (±SD). Maternal age was 29.3 (4.3) years and mean pregnancy BMI was approximately 24 (±4.3) kg/m^2. There were 55% primiparous mothers. Most of the mothers reported never to have smoked during pregnancy while 14% and 12% respectively smoked during pregnancy. About 15% of the mothers had a low degree of education while about half of them obtained a college or university degree. The apparent temperature during the last week of the pregnancy was 9.3 and 9.4 degrees Celsius and the average concentration of $PM_{2.5}$ during the course of the entire pregnancy was ca. 12.5 µg/m^3 (Fig. S1) with the lowest values in the first trimester and the highest values in the last trimester.

Table 1.
Population characteristics and perinatal factors at sampling.

Characteristics	Study-Population (n = 187)		Study-Population (n = 176)	
Newborns				
Girls, n	90	(48%)	84	(48%)
Birthweight, grams	3401	± 476	3409	±462
Western-European, n	168	(90%)	160	(91%)
Gestational age, weeks	39.1	± 1.7	39.1	±1.6
Maternal				
Age, years	29.3	± 4.3	29.3	±4.3
Pre-pregnancy BMI, kg/m^2	24	± 4.3	24.1	±4.3
Education				
Low, n	28	(15%)	27	(15%)
Middle, n	64	(34%)	56	(32%)
High, n	95	(51%)	93	(53%)
Self-reported smoking status				
Smoking during pregnancy, n	26	(14%)	22	(12%)
Parity				
1, n	102	(55%)	96	(55%)
≥2, n	85	(45%)	80	(45%)
Period of conception				
January-March	43	(23%)	42	(24%)
April-June	39	(21%)	36	(20%)
July-September	76	(41%)	69	(39%)
October-December	29	(16%)	29	(16%)
Apparent Temperature, °C	9.3	[-0.08-17.03]	9.38	[-0.1-17.12]
Air pollution estimates (µg/m^3)				
$PM_{2.5}$ during entire pregnancy	12.53	[9.58-15.14]	12.49	[9.5-15.12]
$PM_{2.5}$ during first trimester	11.34	[7.74-15.46]	11.40	[7.78-15.5]
$PM_{2.5}$ during second trimester	12.64	[7.91-17.7]	12.58	[7.9-17.75]
$PM_{2.5}$ during third trimester	13.63	[8.32-19.63]	13.49	[8.26-19.59]

The numbers represent counts (percentages) for categorical and means ± standard deviation for continuous variables. Air pollution estimates are reported as mean [5th–95th centile].

3.2 Evaluation of covariates

The first 10 principal components of the PCA explained ca. 43% of the variance in the methylation (Beta-values), the first principal component explained ca. 17% (Figure S2). Subsequent multiple linear mixed model regression between the principal components against the different variables revealed a strong influence of gestational age on the outcome of the analysis (Fig. 3). Principal component 4 and gestational age showed a strong positive correlation with a *p*-value < 0.0001. Additionally, the principal components 3, 8, and 9 showed negative correlations with gestational age ($p < 0.05$). These findings were confirmed in the analysis of the *QQ* plots which showed much lower *p*-values for the regression with gestational age as experimental variable then expected (Fig. S3 G). Other variables that showed significant correlations with the principal components were birthweight, conception in April-June, pre-gestational maternal BMI, multiparity, apparent temperature during the last week of pregnancy and mediocre and high degree of maternal education. However, these variables showed no clear deviation in the QQ plots (Fig. S3 A-L) except for birthweight which was not included in model 1 because of the bi-directional character of association between birthweight and methylation status. For the top 20 CpGs associated with $PM_{2.5}$ after multiple linear mixed regression with a basic model the lowest *p*-values were generated when adding the variable apparent temperature in the last week of the pregnancy to the model except for CpG site cg03054491 were the highest *p*-value was produced (Figs. S4 A-T).

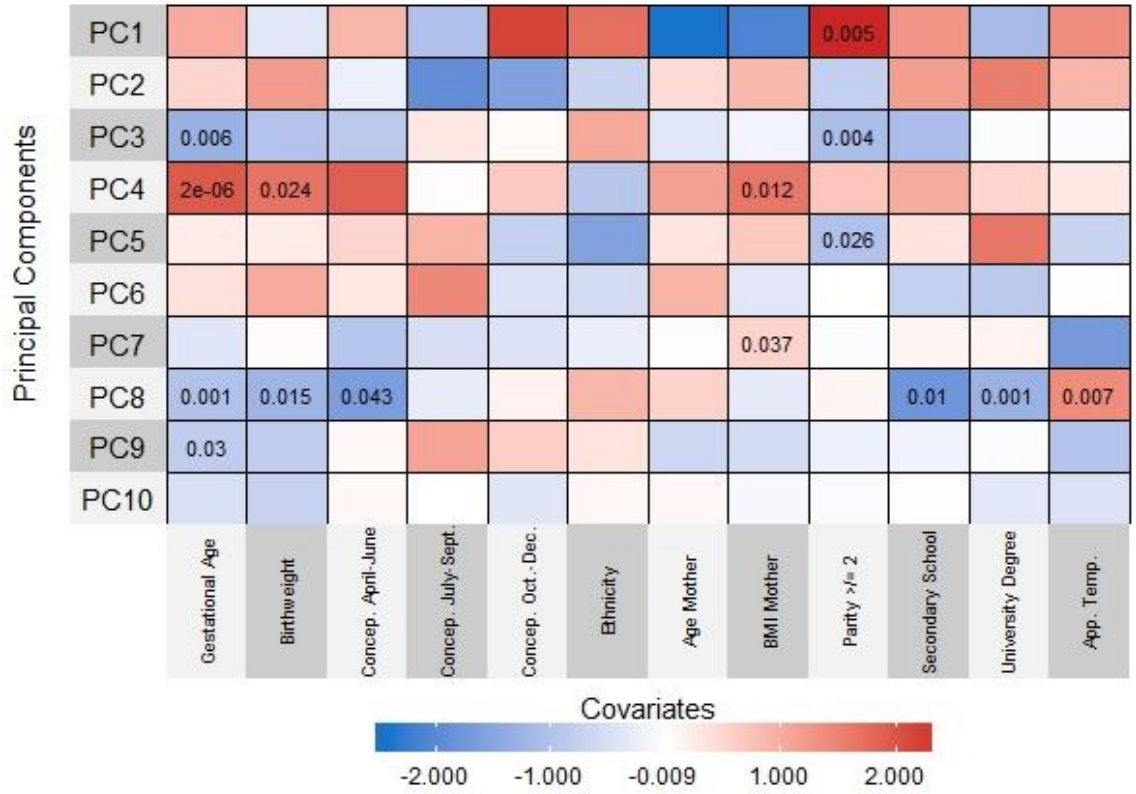


Figure 3. Beta-coefficients and *p*-values for the correlation between the top 10 principal components and the different covariates. Shades of red represent positive correlations and shades of blue negative correlations. Significant correlations ($p < 0.05$) are shown within the corresponding square.

3.3 Association between mtDNA content and $PM_{2.5}$

mtDNA content was negatively associated with conception in July-September (β = -0.16 ± 0.034, p = 2.97e-06) and conception in October-December (β = -0.085± 0.042, p = 0.041) compared to the reference period January-March. After adjusting for apparent temperature during the last week of pregnancy conception in July-September remained significant (β = -0.16 ± 0.034, p = 6.77e-06). None of the associations with other variables that may influence the association between $PM_{2.5}$ exposure and mtDNA content were significant ($p \leq 0.10$) in a multiple linear regression for model 1.

In an unadjusted model mtDNA content and $PM_{2.5}$ were negatively correlated [r=-0.35, p<0.0001 (Figure 4)]. mtDNA content was negatively associated with in utero $PM_{2.5}$ exposure in model 1 (β = -0.038 ± 0.0076, p < 0.0001), adjusting for gestational age, newborn's sex and smoking during pregnancy and number of platelets. After adjusting additionally with the other variables of model 2 (birthweight, ethnicity, parity, season of conception, maternal education, maternal age, maternal pre-pregnancy BMI, apparent temperature during the last week of pregnancy) mtDNA content remained significantly associated with $PM_{2.5}$ (β = -0.021 ± 0.01, p = 0.041). Because the distribution of data from gestational age was skewed, a sensitivity analysis was performed excluding newborns with gestational age of < 37 weeks (n=13). This did not change the association itself (β = -0.038 and -0.021 for model 1 and model 2 respectively) but had influence on the p-value resulting in a loss of significance in model 2 (p = 0.072).

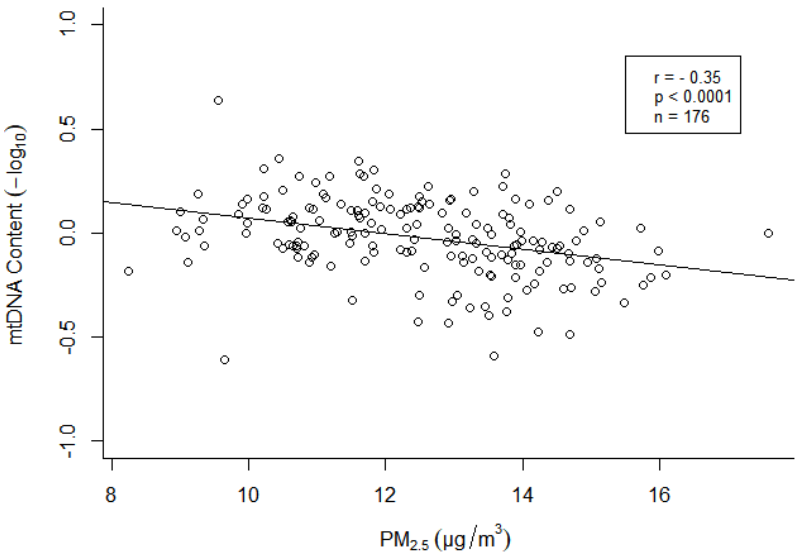


Figure 4. Correlation between *in utero* $PM_{2.5}$ exposure during the entire pregnancy and mtDNA content

When investigated per trimester of the pregnancy $PM_{2.5}$ exposure was not significantly associated with mtDNA content for any of three trimesters in model 2 (Fig. 5). In model 1 only the second trimester was significantly associated to mtDNA content [p < 0.0001, 95% confidence interval (CI): -0.029, -0.012].

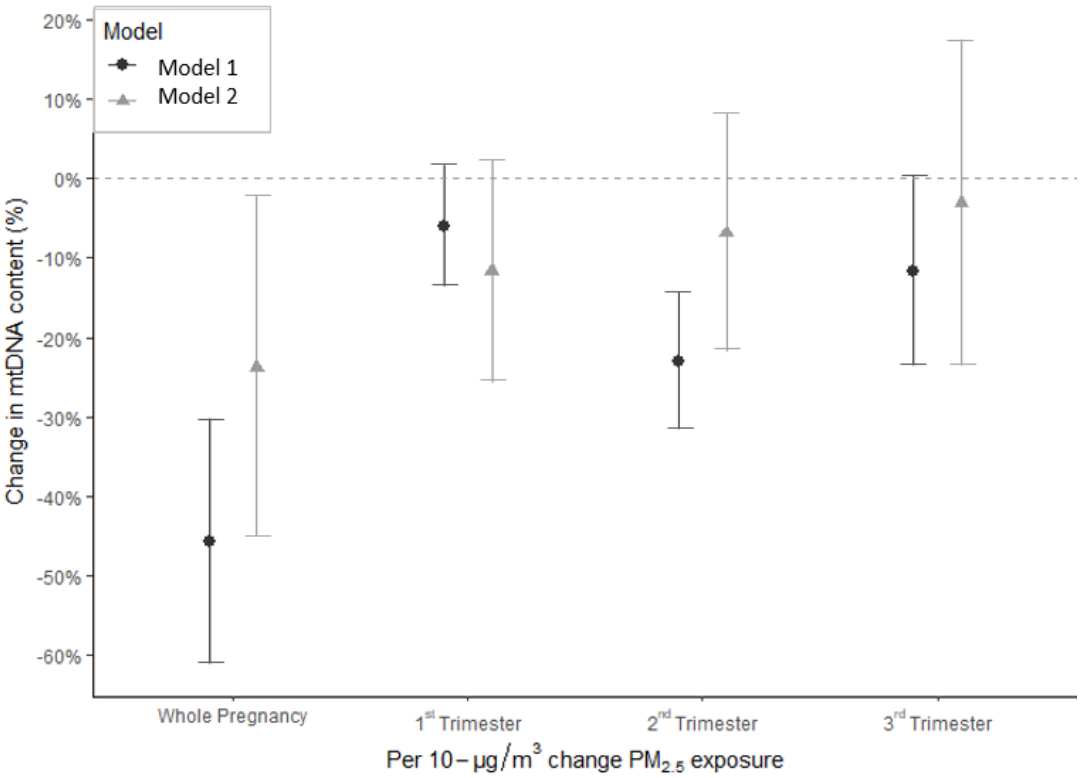


Figure 5. The percent change (95% CI) in mtDNA content for each 10-μg/m^3 increase of $PM_{2.5}$ exposure during specific windows of pregnancy.

3.4 Epigenome-wide association study of the association of individual CpG sites with $PM_{2.5}$ and mtDNA content

The epigenome-wide association study of *in utero* $PM_{2.5}$ exposure with model 1 revealed six differentially methylated CpGs with a *q*-value < 0.2, of which five were annotated (Table 2). All six significant CpG sites showed hypermethylation. The CpG site with the highest *ß*-coefficient and the most significant association with $PM_{2.5}$ (*ß*=0.00568; *p* = 8.57e-07; *q* = 0.183) was located on the *VWA1* gene. Of the individual CpG sites associated with mtDNA content none reached FDR significance (*q* < 0.2).

Table 2.
CpGs significantly associated with *in utero* $PM_{2.5}$ exposure in model 1 (*q*-value <0.2)

CpG	Gene	Position	Localization on CGI	Localization on Gene	*ß*	SE	*p*-Value	*q*-Value
cg09843049	*VWA1*	chr1:1373316	North Shore	Body, 3'UTR	0.00568	0.00111	8.57E-07	0.183
cg07452728	*PIN1*	chr19:9960079	Open Sea	3'UTR	0.00133	0.00026	9.80E-07	0.183
cg26140182	*PSG11*	chr19:43530536	Open Sea	5'UTR, 1stExon	0.00424	0.00085	1.43E-06	0.183
cg26834192	*AGPAT4*	chr6:161561031	Open Sea	Body	0.00315	0.00064	2.25E-06	0.183
cg03804903	*HNRNPM*	chr19:8551040	Island	Body	0.0024	0.00049	2.25E-06	0.183
cg05600804	—	chrX:153623245	North Shelf	—	0.00304	0.00062	2.26E-06	0.183

CpGs sites significant after FDR correction for multiple testing (*q*-value <0.2) are shown. The number of subjects included in model 1 was 187. *ß* value (0–1 scale) represents the difference in methylation for every unit (μg/m^3) increase of air pollutant. UTR = untranslated region. Column headers: Gene = UCSC annotated gene; Position = chromosome and chromosomal position; Localization on CGI= UCSC gene region feature category; Localization on CGI = UCSC relation to CpG islands; *ß* = regression coefficient; SE = standard error for regression coefficient.

3.4.1 Association of the top 20 individual CpG sites from the EWAS with $PM_{2.5}$ and mtDNA content

Among the top 20 of the CpG sites that are highest ranked in terms of *p*-value in the association with $PM_{2.5}$ (Table 3) 11 CpG sites showed a positive association. Sixteen of the 20 CpG sites were annotated. Model 2 applied to the top 20 CpG sites revealed a CpG site annotated to the *UTP20* gene to have the highest significance of association (*p*= 3,98E-06, *q*= 7,95E-05) and being hypomethylated (*ß* = -0,00159).

Table 3.
Association of the top 20 individual CpG sites highest ranked in terms of *p*-value from the EWAS with $PM_{2.5}$ in model 1 and model 2.

CpG	Gene	Model 1				Model 2			
		ß	SE	p-value	q-value	ß	SE	p-value	q-value
cg09843049	VWA1	0.00568	0.00111	8.57E-07	7.52E-06	0.00522	0.00150	6.39E-04	0.00116
cg07452728	PIN1	0.00133	0.00026	9.80E-07	7.52E-06	0.00149	0.00037	8.46E-05	0.00026
cg26140182	PSG11	0.00424	0.00085	1.43E-06	7.52E-06	0.00364	0.00115	1.84E-03	0.00246
cg26834192	AGPAT4	0.00315	0.00064	2.25E-06	7.52E-06	0.00302	0.00075	9.20E-05	0.00026
cg03804903	HNRNPM	0.00240	0.00049	2.25E-06	7.52E-06	0.00213	0.00065	1.32E-03	0.00219
cg05600804	—	0.00304	0.00062	2.26E-06	7.52E-06	0.00304	0.00083	3.52E-04	0.00078
cg14173815	RPL23A, SNORD42A	0.00362	0.00075	3.20E-06	8.45E-06	0.00322	0.00100	1.63E-03	0.00242
cg00308185	BRDT	0.00330	0.00069	3.38E-06	8.45E-06	0.00367	0.00091	8.63E-05	0.00026
cg05534865	—	0.00328	0.00069	4.34E-06	9.64E-06	0.00351	0.00094	2.50E-04	0.00063
cg02930037	RPN2, C20orf132	-0.00194	0.00042	7.39E-06	1.21E-05	-0.00221	0.00050	1.76E-05	0.00018
cg08316775	TPRKB	-0.00105	0.00023	7.44E-06	1.21E-05	-0.00124	0.00029	3.97E-05	0.00024
cg16036471	—	-0.00304	0.00066	7.54E-06	1.21E-05	-0.00293	0.00092	1.69E-03	0.00242
cg17448109	RNF115	0.00308	0.00067	8.45E-06	1.21E-05	0.00285	0.00092	2.34E-03	0.00292
cg27327785	RING1	-0.00180	0.00039	8.50E-06	1.21E-05	-0.00126	0.00053	1.94E-02	0.01937
cg14554618	NAIF1, SLC25A25	0.00236	0.00052	1.12E-05	1.43E-05	0.00227	0.00064	5.53E-04	0.00111
cg04283429	—	-0.00254	0.00056	1.14E-05	1.43E-05	-0.00278	0.00067	4.80E-05	0.00024
cg09842285	UNKL	-0.00089	0.00020	1.26E-05	1.48E-05	-0.00081	0.00027	3.41E-03	0.00402
cg03054491	UTP20	-0.00129	0.00029	1.53E-05	1.70E-05	-0.00157	0.00033	5.37E-06	0.00011
cg04717240	CINP, TECPR2	-0.00189	0.00043	1.71E-05	1.75E-05	-0.00171	0.00060	5.27E-03	0.00585
cg16892812	—	-0.00377	0.00085	1.75E-05	1.75E-05	-0.00313	0.00118	8.87E-03	0.00934

The number of subjects included in the two models is n = 187. *β* value (0–1 scale) represents the difference in methylation for every unit ($\mu g/m^3$) increase of $PM_{2.5}$. Column headers: Gene = UCSC annotated gene; Position = chromosome and chromosomal position; Location= UCSC gene region feature category; β = regression coefficient; SE = standard error for regression coefficient.

Of the top 20 highest ranked CpGs in the association with mtDNA content 15 showed an inverse association (Table 4). The CpG site with the lowest *p*-value in the association with mtDNA content (*p* = 1,60E-06; *q* = 0,465) employing model 1 was annotated to *NOTCH4*. In model 2 a CpG site annotated to *GNA12* showed the lowest *p*-value (*p* = 2,88E-06; *q* = 3,02E-05). Three genes were

represented by multiple CpGs, *NOTCH4* and *SIK3* by two different CpGs each and *TRIM10* by four different CpGs.

Table 4.
Association of the top 20 individual CpG sites highest ranked in terms of *p*-value from the EWAS with mtDNA content in model 1 and model 2.

CpG	Gene	Model 1				Model 2			
		ß	SE	p-value	q-value	ß	SE	p-value	q-value
cg27597473	NOTCH4	-0.0278	0.0056	1.60E-06	1.92E-05	-0.0264	0.0058	1.10E-05	3.65E-05
cg24847621	TRIM10	-0.1047	0.0213	2.13E-06	1.92E-05	-0.0967	0.0238	7.76E-05	0.00015
cg23095517	SOX13	-0.0299	0.0062	3.57E-06	1.92E-05	-0.0223	0.0071	0.00193	0.00193
cg16340918	KIF11	-0.0121	0.0026	5.37E-06	1.92E-05	-0.0129	0.0028	8.57E-06	3.43E-05
cg10283879	—	-0.0420	0.0090	6.66E-06	1.92E-05	-0.0446	0.0096	7.45E-06	3.43E-05
cg02606808	MAP1B	-0.0112	0.0024	7.09E-06	1.92E-05	-0.0113	0.0027	4.11E-05	0.0001
cg16688681	GNA12	0.1149	0.0248	7.66E-06	1.92E-05	0.1365	0.0283	3.37E-06	3.37E-05
cg26611723	—	0.0141	0.0030	7.66E-06	1.92E-05	0.0122	0.0034	0.00049	0.00054
cg07873154	TRIM10	-0.1022	0.0222	8.62E-06	1.92E-05	-0.0998	0.0247	8.29E-05	0.00015
cg11857238	TRIM10	-0.1261	0.0281	1.35E-05	2.51E-05	-0.1170	0.0314	0.00028	0.00037
cg05845236	PCDH9	-0.0161	0.0036	1.46E-05	2.51E-05	-0.0184	0.0038	4.02E-06	3.37E-05
cg08094206	TRIM10	-0.0977	0.0219	1.51E-05	2.51E-05	-0.0878	0.0245	0.00045	0.00052
cg24597131	KIAA1026	-0.0217	0.0049	1.89E-05	2.68E-05	-0.0229	0.0054	3.42E-05	9.78E-05
cg09681286	ABCA3	0.0282	0.0064	1.94E-05	2.68E-05	0.0266	0.0068	0.00014	0.00021
cg13474520	FKTN	-0.0134	0.0031	2.01E-05	2.68E-05	-0.0152	0.0032	5.05E-06	3.37E-05
cg23684410	SIK3	0.0505	0.0115	2.17E-05	2.71E-05	0.0527	0.0126	5.06E-05	0.00011
cg17160660	MYC	0.0588	0.0136	2.71E-05	3.13E-05	0.0468	0.0148	0.00186	0.00193
cg01822785	DDX4	-0.0311	0.0072	2.91E-05	3.13E-05	-0.0281	0.0076	0.00031	0.00039
cg18432864	SIK3	-0.0301	0.0070	2.97E-05	3.13E-05	-0.0297	0.0074	9.86E-05	0.00016
cg04233421	NOTCH4	-0.0472	0.0110	3.18E-05	3.18E-05	-0.0427	0.0114	0.00026	0.00037

The number of subjects included in the two models is n = 176. *β* value (0–1 scale) represents the difference in methylation for a 1% change in mtDNA content relative to its reference genes. Column headers: Gene = UCSC annotated gene; Position = chromosome and chromosomal position; Location= UCSC gene region feature category; β = regression coefficient; SE = standard error for regression coefficient.

3.4.2 Association of the top 20 individual CpG sites from the EWAS with $PM_{2.5}$ per trimester of pregnancy

When examined per trimester, the most significant associations could be found for the third trimester of model 1 (Fig. 6). Six CpG sites showed a significant hypermethylation for the third trimester, four for the second trimester, and only one for the first trimester of model 1 (significance-level $p = (\log_{10}(0.05/20) = \pm 2.6)$ while four, two and one where significantly hypomethylated in the third, second and first trimester respectively. For model 2 three CpG sites were significantly hypermethylated for the third trimester and two for the second trimester. Additionally, three CpG sites were hypomethylated in the third trimester and one in the second trimester in model 2. There were no significant associations for model 2 in the first trimester.

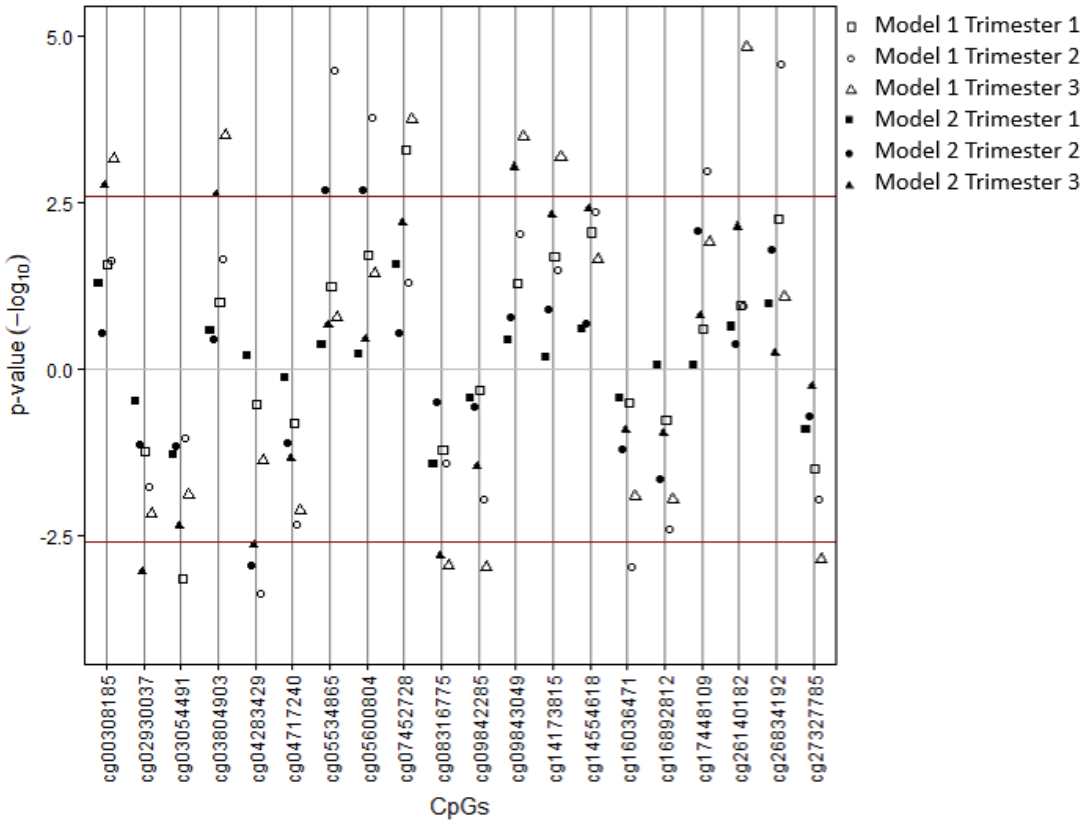


Figure 6. Log_{10}-transformed *p*-values of the top 20 individual CpG sites most significantly associated with $PM_{2.5}$ for all three trimesters in model 1 and model 2. For positive associations the corresponding *p*-values are plotted above the zero y-intercept and for negative ones below. The two red y-intercepts at +/- 2.6 mark the significance level.

3.5 Gene expression and pathway analyses

Within the matrix of 20 × 29,164 = 583,280 CpG–transcript pairs we identified 222 significant pairs, corresponding to 188 unique genes, associated with $PM_{2.5}$, and 832 significant pairs, corresponding to 447 unique genes, for the association with mtDNA content. Gene enrichment analyses performed with the DAVID online tool, applying Reactome, BioCarta & KEGG pathway mapping and gene-disease association, identified several significantly enriched pathways. Pathways were selected with a minimum overlap of 5 genes with the input list and an EASE score below 0.01 and considered significant based on a Bonferroni 5% significance level. We found 23 enriched pathways/gene-disease-associations for $PM_{2.5}$ and 17 for mtDNA content with an overlap of 13 common pathways/gene-disease-associations (Table 5). The most significant pathway in both cases was 'Resolution of Sister Chromatid Cohesion'. The only common gene-disease-associations was for systemic lupus erythematosus. Among the pathways significant for both associations several were associated with histone modification, mitosis and the cell cycle. Pathways only present in the association with mtDNA content were designated to heme synthesis, ROS production in phagocytes and the cross-presentation of particulate exogenous antigens in phagosomes (Table S4). The highest ranking pathway showed involvement in the regulation of the actin cytoskeleton. The pathways exclusively found in association with $PM_{2.5}$ exposure were often involved in DNA damage and repair, telomeres, and senescence. The highest ranking one was related to the transcription process (Table S5).

Table 5.
Pathways and gene-disease-associations commonly found for the association with $PM_{2.5}$ and mtDNA content

Database	Pathway	Term	Association with $PM_{2.5}$			Association with mtDNA content		
			Bonferroni-p-value (5%)	Count %	Fold Enrichment	Bonferroni-p-value (5%)	Count %	Fold Enrichment
Reactome	R-HSA-2500257	Resolution of Sister Chromatid Cohesion	2.75E-13	10.11	13.44	5,45E-10	5.15	6.85
Reactome	R-HSA-68877	Mitotic prometaphase	3.20E-09	7.98	11.89	2,84E-07	4.25	6.34
Reactome	R-HSA-2467813	Separation of sister chromatids	5.28E-08	9.04	7.87	2,67E-06	5.15	4.48
Reactome	R-HSA-156711	Polo-like kinase mediated events	0.0049	2.66	26.75	0.001	1.57	15.78
Reactome	R-HSA-3214858	RMTs methylate histone arginines	2.58E-11	7.98	16.68	0.003	2.68	5.61
Reactome	R-HSA-3214815	HDACs deacetylate histones	4.56E-10	7.98	13.66	0.0038	2.9	4.98
KEGG	hsa05322	Systemic lupus erythematosus	2.26E-10	8.51	11.79	0.0041	3.36	3.97
Reactome	R-HSA-69273	Cyclin A/B1/B2 associated events during G2/M transition	0.018	2.66	19.46	0.0083	1.57	11.46
Reactome	R-HSA-2565942	Regulation of PLK1 Activity at G2/M Transition	0.012	4.26	7.78	0.01	2.68	4.91
Reactome	R-HSA-2299718	Deposition of new CENPA-containing nucleosomes at the centromere	6.49E-09	6.91	15.04	0.013	2.46	5.35
Reactome	R-HSA-606279	Condensation of Prophase Chromosomes	1.44E-11	7.98	17.35	0.013	2.46	5.35
Reactome	R-HSA-983189	Factors involved in megakaryocyte development and platelet production	9.94E-05	4.79	12.23	0.02	2.24	5.72
KEGG	hsa04110	Cell cycle	3.43E-04	5.32	7.96	0.33	2.24	5.71

3.6 Epigenome-wide association study for the association of differentially methylated regions with $PM_{2.5}$ and mtDNA content

3.6.1 DMRcate

We identified 317 DMRs, of which 70 were significant with a Stouffer z-value < 0.2, for the association with $PM_{2.5}$ using the DMRcate algorithm (Table S6 for all significant DMRs and Table 6 for the top 20 highest ranked according to *p*-value). The DMR showing the highest significance of association expressed by the Stouffer-value was located inside an exon of the gene *RPTOR* on position 78863570-78866579 of chromosome 17 (Fig. 7 A). The mean increase in DNA methylation for the CpG sites contained in this DMR was 6.8% for each 10-µg/m^3 increase of $PM_{2.5}$. Two of the 12 CpG sites that constituted this DMR showed an FDR-value below the threshold of 0.2. All 12 individual CpG sites were hypermethylated and located on a CpG island or its shores or shelves (Table S7). The highest effect on methylation was found for the DMR annotated to *FLJ13224* with a mean increase in DNA methylation of 16.7% for each 10-µg/m^3 increase of $PM_{2.5}$ (Stouffer z-value = 0,1482). Two CpG sites on DMRs identified by DMRcate for the association with $PM_{2.5}$ were also among the six CpG sites significantly associated with $PM_{2.5}$ in the EWAS. One of them, cg09843049, was annotated to *VWA1* encoding for the "*von Willebrand* factor" (vWf) protein (Fig. 7 B). Although DMRcate does not require

all CpGs within a DMR to have the same direction of association with the independent variable, all four CpG sites constituting this DMR were hypermethylated and located on a CpG island shore (Table S8). The other DMR with a CpG site already identified by the EWAS was located on *PSG11* and constituted of two CpG sites both hypermethylated and not located within a CpG island (Table S9). Two of the four CpG sites were below the significance threshold of FDR-value <0.2 and therefore taken into account for the allocation of DMRs. The vast majority of the significant DMRs identified by DMRcate for the association with $PM_{2.5}$ (89%) showed hypermethylation when evaluating the mean beta-fold change. Only eight of the 70 DMRs were hypomethylated. On the other hand, for the association with mtDNA content no associations were found with the DMRcate approach. Because in this case no output is generated by DMRcate, a comparison of the outcomes for the association with $PM_{2.5}$ and mtDNA content was not possible.

Table 6.
Top 20 of the differentially methylated regions highest ranked in terms of Stouffer z-score identified by the DMRcate algorithm for the association with $PM_{2.5}$ in model 1.

Gene	Coordinates	Location	no.cpgs	meanbetafc	Stouffer
RPTOR	chr17:78863570-78866579	covers exon(s)	12	0.0068	0.0113
RNF126	chr19:649039-649464	covers exon(s)	3	0.0021	0.0411
UBE2MP1	chr16:34456419-34456429	upstream	2	0.0044	0.0480
VWA1	chr1:1373044-1374310	inside intron	4	0.0034	0.0797
LOC441242	chr7:64894934-64895418	downstream	3	0.0031	0.0802
SNORA71B	chr20:37054880-37056075	promoter	4	0.0021	0.0887
NDRG2	chr14:21526267-21526351	inside intron	2	0.0033	0.0927
AGPAT5	chr8:6470040-6470839	upstream	4	0.0020	0.0959
KDM5B	chr1:202778443-202779497	promoter	5	-0.0089	0.0988
CCHCR1	chr6:31116343-31116408	overlaps exon	2	0.0019	0.1019
NFIB	chr9:14314066-14314158	inside intron	2	-0.0029	0.1035
ZNF767P	chr7:149461522-149461763	upstream	2	0.0022	0.1106
TRIP10	chr19:6670865-6671045	upstream	2	0.0023	0.1115
ATP5EP2	chr13:28519319-28519388	overlaps 5'	2	0.0031	0.1149
USP28	chr11:114127924-114127960	upstream	2	0.0024	0.1162
HDAC4	chr2:240196769-240197131	inside intron	4	0.0024	0.1162
PEX11G	chr19:7676731-7676786	upstream	2	0.0026	0.1225
TELO2	chr16:1444138-1444162	upstream	2	0.0023	0.1231
CFAP46	chr10:134699444-134699744	downstream	3	0.0017	0.1238
ADGRB1	chr8:143545940-143545949	inside exon	2	0.0042	0.1308

Column headers: Gene = UCSC annotated gene; Position = chromosome and chromosomal position; Location= UCSC gene region feature category; no.cpgs = number of CpGs constituting the differentially methylated region; meanbetafc = mean ß fold change; Stouffer = Stouffer's z-score.

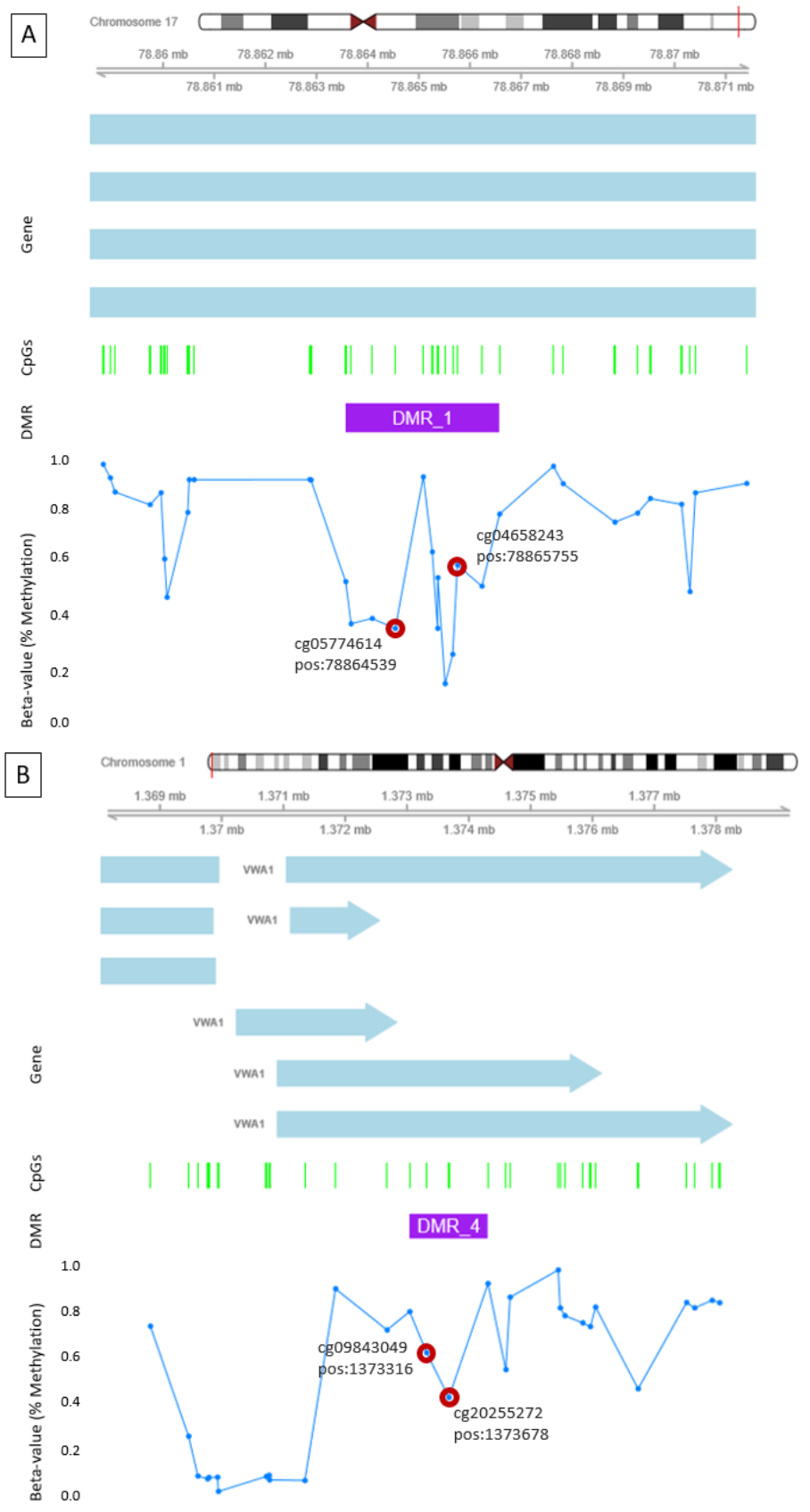


Figure 7. (A) Location of the differentially methylated region (DMR) most significantly associated with $PM_{2.5}$ identified by the DMRcate algorithm and (B) of the DMR annotated to the gene *VWA1*. All CpG sites, also individually non-significant ones, within the identified region are shown together with their Beta-value (% of methylation) and position on the chromosome. The CpG sites which are individually significant at a FDR < 0.2 level are encircled in red and shown with their exact position on the chromosome.

3.6.2 Bumphunter

When using the bumphunter algorithm none of the 810 of the DMR candidate regions found for the association with $PM_{2.5}$ were significant at a 0.2 FWER level. For the association with mtDNA content two DMR candidate regions, both of which were hypermethylated, were found significant ($q < 0.2$) after correction for FWER. These DMRs were annotated to the genes *STIL* and *SPRN* respectively. Even though the overall direction of effect in case of the DMR annotated to *STIL* was positive, the two CpG sites constituting the DMR showed an opposite direction for both associations (Table S10). For the DMR located on the gene body of SPRN all 11 CpG sites were hypermethylated and within a CpG island or its shore (Table S11).

In the top 20 of the candidate DMRs highest ranked in terms of FWER most for the association with $PM_{2.5,}$ eight showed a positive association (average hypermethylation) and 12 a negative association (average hypomethylation) (Table S12) while for the association with mtDNA content 11 were hypermethylated and nine hypomethylated (Table S13). Additionally, an overlap of four genes was found between the genes annotated to the 20 top ranked candidate DMRs found by bumphunter for the association with $PM_{2.5}$ and mtDNA content respectively (Table 7; Fig. S6). These four common genes included: *HLA-DRB6*, *CCKBR*, *CPT1B* and *STIL,* with the latter being significant and ranked first for the association with mtDNA content and second for the association with $PM_{2.5}$. For the DMRs annotated to *STIL* and *CCKBR* the overall direction of effect was the same (hypermethylation) for both associations but for *HLA-DRB6* and *CPT1B* it showed an opposite direction. For the association with mtDNA content four CpG sites less than for the association with $PM_{2.5}$ were constituting the DMR annotated to *HLA-DRB6* (Table S14). Here all CpG sites were hypomethylated for the association with mtDNA content and all but one CpG site were hypermethylated for the association with $PM_{2.5}$. Furthermore, there was also one CpG site less found for the association with mtDNA content compared to $PM_{2.5}$ in the DMRs annotated to *CPT1B* and *CCKBR* respectively with all CpG sites being affected in the same direction per DMR and association (Tables S15 and S16). None of the top 20 ranked candidate DMRs for the associated with mtDNA content were present in the top 20 of individual CpG sites associated with mtDNA content in the EWAS. Furthermore, there was no overlap between the 70 significant DMRs found by DMRcate for the association with $PM_{2.5}$ and the top 20 ranked candidate DMRs for the associated with mtDNA content identified by bumphunter.

Table 7.
Overlap of differentially methylated regions (DMRs) in the top 20 highest ranked according to FWER associated with $PM_{2.5}$ or mtDNA content identified by the bumphunter algorithm.

Gene	Coordinates	Location	$PM_{2.5}$			mtDNA content		
			no.cpgs	meanbetafc	FWER	no.cpgs	meanbetafc	FWER
STIL	chr1:47800167-47800409	upstream	2	0.022	0.35	2	0.504	0.08
HLA-DRB6	chr6:32551749-32552246	overlaps 5'	14	0.006	0.57	10	-0.050	1.00
CCKBR	chr11:6291339-6292896	covers	12	0.005	0.84	11	0.037	1.00
CPT1B	chr22:51016501-51017151	overlaps 5'	13	-0.004	0.95	12	0.035	1.00

Column headers: Gene = UCSC annotated gene; Position = chromosome and chromosomal position; Location= UCSC gene region feature category; no.cpgs = number of CpGs constituting the differentially methylated region; meanbetafc = mean ß fold change; FWER = Family wise error rate.

3.7 Mediation analysis

Two of the six CpG sites significantly associated with $PM_{2.5}$ in the EWAS also showed a *p*-value < 0.05 for the association with mtDNA content in a multiple linear mixed regression model only adjusted for batch effects. Mediation analysis on these CpG sites, cg03804903 and cg26834192, showed that mtDNA content mediated 33.7% (DE = 0.68 %, p=0.08; IE = −0.35 %, p=0.03) of the association between prenatal $PM_{2.5}$ exposure during the entire pregnancy and methylation of cg03804903 in a model with partial adjustment for covariates (Fig. 8). In model 2 applied for cg03804903 the IE, representing the effect of mediation was not significant (*p*-value = 0.11) (Fig. S6). For cg26834192 the mediation of mtDNA content in the association between prenatal $PM_{2.5}$ exposure and methylation status was neither significant in model 1 nor in model 2 (Figs. S7).

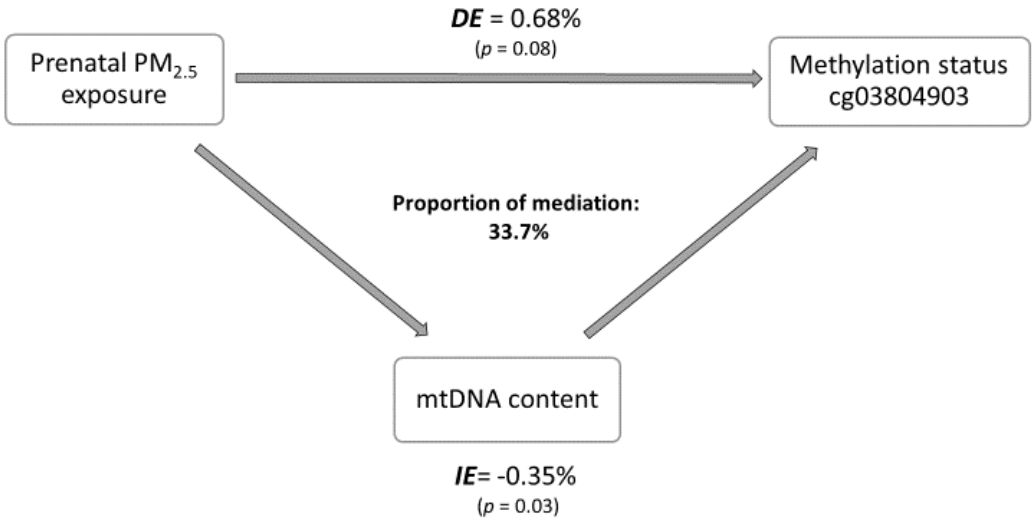


Figure 8. Estimated proportion of the association between a 10 µg/m^3 increment in $PM_{2.5}$ exposure during the entire pregnancy and methylation status of CpG site cg03804903 mediated through mtDNA content. Additionally, the estimates of the indirect effects (IE), the estimates of the direct effect (DE), and the proportion of mediation (IE/DE+IE) are shown. The model was adjusted for chip, position on the chip, white blood cell composition, platelet count, maternal smoking status and newborns' sex and gestational age.

4 Discussion

Prenatal exposure to ambient air pollution could affect birth outcomes and may have the potential to evoke adverse health outcomes later in life. With this study we provided evidence on different molecular levels for a *in utero* effect of $PM_{2.5}$ exposure on cord blood mtDNA content and epigenome-wide DNA methylation representing biomarkers of early effects on a genetic and epigenetic level, respectively. Furthermore, we found an overlap between (i) pathways for transcripts corresponding to epigenetic signatures of $PM_{2.5}$ exposure and mtDNA content and (ii) DMRs for both associations indicating possible interactions. Finally, we could demonstrate a partial mediation of $PM_{2.5}$ exposure effects on the methylation status of a CpG site by mtDNA content.

In a model adjusted for platelet count, maternal smoking status, sex and gestational age of the child, we found prenatal $PM_{2.5}$ exposure to be inversely associated with mtDNA content. This association persisted after additional adjustment for birthweight, ethnicity, parity, season of conception, maternal education, age, pre-pregnancy BMI, and apparent temperature during the last week of pregnancy but not after excluding pre-term births from the study. Our observation is in agreement with previous studies that also documented an inverse association between *in utero* $PM_{2.5}$ exposure and mtDNA content in cord blood (51) and placenta (50). Furthermore, a study on the effects of PM on the telomere–mitochondrial axis of aging in elderly participants (82) also showed a negative effect of air pollution on mtDNA content. Contrary to the aforementioned studies of prenatal $PM_{2.5}$ exposure which found the strongest effect on mtDNA content for the third trimester of pregnancy, we found the strongest effect during the second trimester of pregnancy. With regard to the study on mtDNA content in placental tissue this difference may be explained by the distinctive functions and energetical demands of this two media. Our observations could be explained by the high vulnerability of mtDNA for PM induced damage. After persistent exposure and continuous attacks, the mtDNA damage exceeds the compensatory capacity of the mitochondria, resulting in a decrease of mtDNA content (47, 48). The consequences for the fetus may include a lower birth weight (83) and a restricted early childhood growth (84) as demonstrated previously in the context of an association with NO_2. In adolescence a reduced mitochondrial content in peripheral leucocytes has been associated with insulin resistance (85) which could indicate a higher susceptibility to certain diseases later in life as a consequence of prenatal $PM_{2.5}$ exposure.

Our study investigated epigenome-wide DNA methylation in association with *in utero* $PM_{2.5}$ exposure and mtDNA content using an untargeted approach, which enabled us to consider a vast number of methylation targets in a unbiased manner. The EWAS revealed six individual CpG sites significantly associated with *in utero* $PM_{2.5}$ exposure, all showing hypermethylation. This is in accordance with findings of a longitudinal epigenome-wide methylation study on the effects of $PM_{2.5}$ species in 646 individuals (58) and a meta-analysis on the effect of PM exposure in three cohorts (86) also reporting predominantly hypermethylation of individual CpG sites.

We identified several novel individual CpG sites in the EWAS not previously associated with air pollution and often linked with cytokine secretion and inflammation. Located on the 5' untranslated region (5'UTR) region and the 1st exon of *PSG11,* cg26140182 was found to be hypermethylated. *PSG11* encodes for human pregnancy-specific glycoproteins (PSGs) which are mainly produced by the placental syncytiotrophoblasts during pregnancy and modulate the innate immune system by inducing dose-dependent secretion of anti-inflammatory cytokines (87). In clinical studies, low serum PSG concentrations during the first trimester of pregnancy and maternal copy-number deletion in *PSG11*

were correlated with adverse pregnancy outcomes including pre-term delivery, small birth weight and preeclampsia (88, 89). Furthermore, the risk of preeclampsia was also found to be increased in relation to prenatal $PM_{2.5}$ exposure (90, 91). $PM_{2.5}$ induced hypermethylation of *PSG11* may result in its transcriptional downregulation with potentially consequences for the embryogenesis linking prenatal $PM_{2.5}$ with adverse pregnancy outcomes. On the gene body of heterogeneous nuclear ribonucleoprotein M (*HNRNPM)* we also found the probe cg03804903 to be hypermethylated. This CpG site was located within a CpG island which may indicate transcriptional influence, but due to the localization on the gene body conclusions cannot be drawn. HNRNPM may initiate a signaling cascades which ultimately leads to the induction of IL-1 alpha, IL-6, IL-10 and TNFα and functions as a mediator of metastasis and the inflammatory response (92). In a study comparing small airway epithelium of 19 nonsmokers and 20 smokers *HNRNPM* was found among the top 25 hypermethylated genes most significantly associated with cigarette smoking (93). In our study, among the top 20 ranked individual CpG sites various others were located on genes associated with the immune system like *NF115* encoding for Ring Finger Protein 115 and *UNKL* encoding for the 'Unkempt Family Like Zinc Finger' protein which are both related to the pathways 'Innate Immune System' and 'Class I MHC mediated antigen processing and presentation'. Others were linked to gene transcription and DNA replication such as *RING1* which acts as a transcriptional repressor throughout development, or *CINP*, a component of the DNA replication complex which interacts with proteins imperative for the replication initiation.

Top significant CpG site associated with *in utero* $PM_{2.5}$ was cg09843049 which is located on an island shore on von Willebrand Factor A Domain Containing 1 (*VWA1)* encoding for the 'von Willebrand factor' (vWf), a blood glycoprotein with several domains capable of binding molecules related to blood coagulation and formation of blood vessels (94). Because of its involvement in thrombotic platelet plug formation and myocardial and cerebrovascular infarctions it is also utilized as a biomarker of risk of chronic cardiovascular events (95). The association of vWf with air pollution, especially $PM_{2.5}$, has already been described extensively in previous studies in adults with or without a disease background. Evidence was found for an upregulation in vWf, significantly associated with a 5-day mean increase in $PM_{2.5}$ in a study of coronary artery disease in Germany (96). Furthermore, a significant relationship was also demonstrated for increased concentrations of vWf with components of in-vehicle $PM_{2.5}$ in a study among highway patrol troopers (97). Comparable observations were made for the association between ambient PM_{10} among diabetics in the 'Atherosclerosis Risk in Communities' study, were a 12.8-$\mu g/m^3$ increase in PM_{10} was associated with a 3.9% higher level of vWf (98). More recently two independent studies (99, 100) described the effects of improved air quality accompanying the period of the Olympic Games in Beijing 2008 on biomarkers of inflammation and compared them with observations before the start of the Olympic Games and after termination of pollution control. Both studies found statistically significant increases of vWf associated with increasing levels of $PM_{2.5}$. This results appear to be in contrast with our finding of hypermethylation of a CpG sites located on an island shore because methylation in this region is classically associated with gene silencing. CpG island's shores describe the 2 kb of sequences flanking a CpG island which demonstrate a higher methylation dynamic than the island itself (66) and whose methylation status is suggested to affect gene expression (101). In this case though the localization of the CpG island was inside the gene body and on the 3'UTR. The function of DNA methylation in the gene body and the 3'UTR is uncertain and associations with gene expression levels cannot generally be inferred from methylation data for this regions (102). While there is a general consensus that methylation in the close proximity of the transcription start site (TSS) blocks transcription initiation,

methylation in the gene body does not always have this effect but might rather stimulate transcription elongation, as considerable positive correlations between active transcription and methylation at the gene body have been reported (103).

Among the top 20 CpG sites most significantly associated with $PM_{2.5}$ we could not find any genes directly involved in the regulation or maintenance of mitochondrial DNA or functions. A reason other than a lack of interaction could be that, with regard to the high number of CpG sites present on the array (n = 485577), the restriction to 20 CpG sites is too small to investigate an overlap.

In analyses using $PM_{2.5}$ exposures during each trimester of the pregnancy positive as well as negative associations were stronger in the last trimester and weaker in the first trimester. The effect of prenatal $PM_{2.5}$ exposure on epigenome-wide cord blood methylation has to our knowledge not yet been investigated. Therefore, a direct comparison with literature for the association in different trimesters is not possible. Elsewhere the effect of $PM_{2.5}$ on the methylation status of CpG sites on nuclear (104) and mitochondrial (50) DNA in placental tissue has been examined. In the first study the associations were most significant in the second trimester, and in the latter in the first trimester. These differences could maybe be explained by the disparity of the investigated CpG sites and underlying gene functions and also by the distinctive purpose and energetical demands of the placenta, which acts as a temporary organ for the accommodation of the fetus.

None of the associations between mtDNA content and individual CpG sites was significant in the EWAS after correcting for multiple testing with a FDR threshold of 0.2. One reason for this could be the imperative to meet the high requirements imposed by the multiple correction in EWASs. Additionally, the sample size of n = 176 may be insufficient to provide the required statistical power.

Strikingly when observing the 20 top hits according to p-values four different probes were located on the same gene, *TRIM10*. The corresponding protein plays a role in terminal differentiation of erythroid cells and its related pathways include 'Cytokine Signaling in Immune system' and 'Innate Immune System'. Because the CpG sites were all located outside a CpG island and on the gene body the consequences of the observed hypomethylation remain unclear. One other probe (cg17160660) found in the 20 top hits according to p-values was located on the *Myc* proto-oncogene encoding for a transcription factor which activates transcription of growth-related genes and plays an important role in the biogenesis of mitochondria. *Myc* expression stimulates the mitochondrial biogenesis in the critical phase of cell-cycle entry by increasing the substrate availability for enzymes necessary to ensure the programmed expansion of mitochondrial content (105). Additionally, it also indirectly controls mitochondrial gene expression by repressing microRNAs regulating the expression of nuclear genes encoding mitochondrial proteins (106). Flow cytometric studies in rat fibroblasts, have previously revealed that *Myc* null cells possess reduced mitochondrial mass and function (105, 107). In our study we detected a hypermethylation of the CpG site located on *Myc* on the north shore of a CpG island within 1500 bp upstream of the TSS. This observation could therefore point to a downregulation of *Myc* expression with adverse consequences for the mtDNA content.

As reported for the top 20 CpG sites most significantly associated with $PM_{2.5}$, also for mtDNA content the significance of association was greater in model 1 than model 2 but with the same direction and strength of effect.

Pathway analysis of transcripts significantly correlated with the top 20 CpG sites from the two EWAS with mtDNA content and *in utero* $PM_{2.5}$ exposure revealed overlap of 13 pathways. These common pathways were mainly related to histone modification and processes found in mitosis and the cell cycle

and were indeed retraceable to already described effects associated with air pollution and mitochondria. In case of the Reactome pathways 'HDACs deacetylate histones' and 'RMTs methylate histone arginines' histone deacetylase (HDAC) and arginine methyltransferases (RMTs) promote the removal of acetyl groups from histones and the methylation of histone bound arginines respectively. In both cases the enzymatic action results in coiling of the chromatin structures and therefore transcriptional silencing (108). Exposure to diesel exhaust particles (109) and cigarette smoke (110) have been reported to decrease the HDAC2 activity and promote acetylation while conversely mitochondrial dysfunction due to the loss of mtDNA has been shown to decrease site-specific histone acetylation marks (111). Mitochondrial function is crucial to provide intermediate metabolites necessary to create and modify histone acetylation and methylation as for instance in the synthesis of S-adenosylmethionine (SAM). On the other hand, mitochondrial dysfunction increases ROS, succinate and fumarate which in turn inhibit histone demethylases and therefore also induce histone hypermethylation (112). Furthermore, among the pathways shared by mtDNA content and exposure to $PM_{2.5}$ we found 'Cyclin A/B1/B2 associated events during G2/M', 'Regulation of PLK1 Activity at G2/M Transition' and 'Cell cycle'. In literature aberrations of the mitotic spindle with disturbance of the organization of microtubules and centrosomes amplification have been observed after exposure to $PM_{2.5}$ (113), and it was suggested that PM may act as a spindle poison, which disturbs microtubules dynamics and ultimately results in cell cycle arrest and chromosomal aberrations (114). In addition, mitochondria are crucial for cell-cycle progression as this process requires a significant amount of energy. Cyclin B1/Cdk1 increases mitochondrial ATP generation by mediating the phosphorylation of mitochondrial substrates and therefore provides cells with sufficient energy for successful G2/M transition and cell-cycle progression (115).

We additionally identified one gene-disease association for $PM_{2.5}$ as well as mtDNA content. 'Systemic lupus erythematosus' (SLE) has also been linked previously with DNA hypomethylation especially of CD4+ T cells (116-118) and with PM exposure and mitochondrial dysfunction. A longitudinal panel study of juvenile-onset SLE patients living in Sao Paulo showed a significant association between the risk of an SLE Disease Activity Index 2000 (SLEDAI-2K) ≥ 8 and increased PM exposure (119). In another study an increased risk of SLEDAI-2K ≥ 8 was associated with a $PM_{2.5}$ 7-day moving average (120). In yet another study the mitochondrial transmembrane potential and production of ROS intermediates were found to be elevated in the SLE group compared with the healthy controls. It is suggested that mitochondrial hyperpolarization causes ATP depletion which ultimately results in cell necrosis, contributing to inflammation in patients with SLE (121, 122). A link between $PM_{2.5}$ and mtDNA content in the context of SLE could possibly be found in adverse effects of $PM_{2.5}$ induced oxidative stress which results in mitochondrial dysfunction and consequently leads to ATP depletion with diminished activation-induced apoptosis in favor of necrosis. Interestingly TRIM21 a member of the superfamily of tripartite motif containing proteins, acts as an autoantigen which is recognized by antibodies of patients with systemic lupus erythematosus (123). TRIM21 and TRIM10 of which the corresponding gene was represented four times in the top 20 of the CpG sites associated with mtDNA content, belong to the same family IV of the TRIM superfamily. Considering the significant protein domain homology between different TRIM proteins especially of the same family it could be interesting to further investigate potential associations between *TRIM10* and SLE with respect to mtDNA content.

We not only assessed the associations between cord blood epigenome-wide DNA methylation with prenatal $PM_{2.5}$ exposure and mtDNA content on the level of individual CpG sites but also on a regional level by applying two different algorithms from Bioconductor packages DMRcate and

bumphunter. This approach provides more insights into the underlying biological processes, gives context to findings of differential methylation on CpGs and also improves statistical power (124). Because of the hypothesis generating nature of the comparison between associations linked with $PM_{2.5}$ and mtDNA, and in analogy with the EWAS, we only employed model 1 for the analysis. For DMRcate we allowed individual CpGs to enter the calculation of the regions at a FDR threshold of 0.2 in order not to exclude potential candidate regions from being reported. However, we did not employ the marginal *p*-value, the lowest *p*-value of an individual CpG site within the DMR as the criterium for significance of the entire DMR but the much stricter Stouffer z-score. Only those DMRs were considered significant with a Stouffer z-value < 0.2. Applying this criteria DMRcate could not identify any DMRs for the association with mtDNA content. For the association with $PM_{2.5}$ differential methylation was observed in 70 DMRs related to histone modification, mitosis and cell cycle progression, mitochondrial biogenesis and the respiratory chain, conform with our findings in the EWAS and pathway analysis. The DMR most significantly associated with *in utero* $PM_{2.5}$ exposure was located on the body of the gene encoding the Regulatory Associated Protein Of MTOR Complex 1 (RPTOR). We found that the CpG sites, spread over a CpG island and its shores and shelves located at an exon of RPTOR, were hypermethylated in subjects with a higher prenatal $PM_{2.5}$ exposure suggesting downregulation of gene expression. The signaling pathway controlled by MTOR Complex 1 (mTORC1) regulates cell growth in response to various parameters such as hormonal signals. RPTOR associates with the eukaryotic initiation factor 4E-binding protein-1 and hereby allows mTORC1 to control mitochondrial activity and biogenesis via targeted promotion of nucleus-encoded mitochondrial protein synthesis (125). One of the functions of mTORC1 is also to phosphorylate and inhibit ATG1 protein kinase which mediates initial steps in autophagy (126). The transcriptional downregulation of RPTOR which renders mTORC1 inactive may therefore have the potential to increase autophagy resulting in decreased mtDNA content. Additionally, RPTOR is capable of promoting mitochondrial uncoupling as shown in a study of (*raptor*$^{ad-/-}$) knockout mice (127). A CpG site, probed by cg08314949 and located on the body of *RPTOR* but not among the 12 CpGs represented by the DMR found in this study was also significantly hypermethylated in association with prenatal NO_2 exposure in a epigenome-wide meta-analysis in cord blood samples (128).

Remarkably, two of the six CpGs found to be significant in the EWAS with $PM_{2.5}$ were also among the 70 significant DMRs identified by DMRcate for the same association (for the function of the encoded proteins and previous literature on the relationship with air pollution see above). In case of *VWA1* a hypermethylation of the CpG sites located in an intron within the gene body were detected. The methylation of introns can have an opposite effect than the methylation found in the promoter and exon region (129). It was suggested that introns can act as insulators, or a novel, not yet identified mechanism could upregulate gene expression through a combination of hypermethylation and structural confirmation of the intron. This conclusion was made after observing high levels of EGR2 in cell lines with increased methylation at CpG islands in intron 1 of *EGR2* and *vice versa* a decreased expression in lines with a hypomethylated intron (129).

The other DMR with a CpG site also identified for a significant association with $PM_{2.5}$ in the EWAS was located on an exon of *PSG11* and found to be significantly hypermethylated. This could indicate a transcriptional downregulation of the corresponding gene product.

Of the 70 significant DMRs identified by DMRcate for the association with $PM_{2.5}$ we further found six with an annotation to genes related to mitochondrial biogenesis and the respiratory chain: *CKMT2*, *MTG1*, *ATP5EP2, NDUFA3* and *NDUFS6*. Likewise, three DMRs with an annotation to genes coding for

enzymes involved in demethylation and deacetylation were identified: *KDM5B, TET1* and *HDAC4. KDM5B* encodes Lysine (K)-specific demethylase 5B, a lysine-specific histone demethylase which plays a role in the transcriptional silencing of specific tumor suppressor genes by demethylating lysine 4 of histone H3. It is upregulated in various cancer cells and involved in genome stability and DNA repair (130). Here, we found a hypomethylation of the DMR located at the promotor of the gene suggesting an upregulation of gene expression. For Tet Methylcytosine Dioxygenase 1 *(TET1),* encoding another enzyme responsible for demethylation, the corresponding DMR, upstream of the gene, was hypermethylated. TET1 actively removes methyl groups independent of DNA replication. It acts by hydroxylation of methyl residues on cytosines, creating 5hmC which leads to demethylation through base excision repair (BER) mechanisms (131). For the hydroxylation of 5mC, TET enzymes depend on oxygen, connecting their activity and consequently DNA methylation with mitochondrial (dys)function (112). Expression of *TET1* has been shown to decrease in dendritic cells after the exposure to $PM_{2.5}$, aggravating the inflammation in allergic rhinitis (132). In another study conducted in 60 truck drivers and 60 office workers in Beijing, China, no correlation between 5hmC and personal measures of $PM_{2.5}$, though one with ambient PM_{10} was documented and attributed to a decrease in TET1 (133). We also found the DMR located in an intron of Histone Deacetylase 4 (*HDAC4)* to be hypermethylated. HDAC4 is an enzyme which is in charge of the deacetylation of lysine residues located on the N-terminal section of the core histones. It 'tags' them for epigenetic repression and is therefore involved in the transcriptional regulation and cell cycle progression. Oxidative stress generated by air pollutants is known to inhibit HDAC activity, consequently enhancing transcription of proinflammatory genes by NF- κB activation (134, 135). Additionally, we found a DMR in an intron sequence of Mitotic Arrest Deficient 1 Like 1 *(MAD1L1)* to be significantly hypermethylated, in accordance with our findings of two common pathways for $PM_{2.5}$ and mtDNA content related to the separation of sister chromatids. The encoded protein is a component of the mitotic spindle-assembly checkpoint and serves to prevent the start of anaphase until the chromosomes are correctly aligned at the metaphase plate. *MAD1L1* is therefore involved in cell cycle control and tumor suppression and contained in the aforementioned pathways of separation and resolution of sister chromatids.

Using bumphunter we were not able to identify significant DMRs for the association with $PM_{2.5}$ at a FWER threshold of 0.2. This could be due to the different algorithms and correction methods for multiple testing applied by the two packages DMRcate and bumphunter. Two DMRs were though significant for the association with mtDNA content: *STIL* and *SPRN*. *STIL* was also among the four DMRs present in both top 20 lists of DMRs associated with $PM_{2.5}$ or mtDNA content. *STIL* encodes for a cytoplasmic protein required for the activation of the mitotic spindle checkpoint by initiating the onset of procentriole formation. It further facilitates chromosome segregation and mitotic progression during cell division and disappears after transition to the G1 phase (136). It relates to the second most significantly enriched common pathway found in this study, 'Mitotic prometaphase'. For the association with $PM_{2.5}$ as well as mtDNA content we observed a hypermethylation of the two CpG sites located at a CpG island upstream of *STIL* which suggests a transcriptional downregulation. The other DMR significantly associated with mtDNA content, was annotated to *SPRN* which encodes a protein possibly protecting the normal cell-surface glycoprotein from conformational alteration into its pathological isoform (137) and has not been described earlier in literature for the association with mtDNA content or $PM_{2.5}$. We found the CpG sites constituting this DMR, located upstream of the gene body and within a CpG island and shores to be hypermethylated. Except *STIL*, three other common genes were found in the top 20 lists of DMRs associated with $PM_{2.5}$ or mtDNA content: *HLA-DRB6*, *CCKBR and CPT1B. HLA-*

DRB6 represents a highly variable pseudogene of the human major histocompatibility complex (MHC) lacking exon 1, promoters and the 3' untranslated region (3'UTR) polyadenylation signal, though its molecules may be expressed on the cell surface (138). The analysis with bumphunter showed a positive association for $PM_{2.5}$ and an inverse one for mtDNA content with the latter accounting for four CpG sites less than the first. The CpG sites were located on an CpG island and shores with the DMR itself overlapping the genes 5'UTR region. With regard to the transcriptional regulation, the impact of $PM_{2.5}$ and mtDNA content on the transcriptional regulation could therefore be of opposite effects with higher $PM_{2.5}$ exposure and lower mtDNA content inducing a transcriptional downregulation of *HLA-DRB6*. The third DMR that showed an overlap was annotated to Carnitine Palmitoyltransferase 1B (*CPT1B*) encoding for an enzyme that controls the rate of the long-chain fatty acid ß-oxidation pathway in the mitochondria of myocytes where it is responsible for the transport of long-chain fatty acyl-CoAs from cytoplasm into mitochondria. Worth mentioning is the fact that transcripts are expressed from the upstream locus including some of the gene's exons (139). Here, we also observed different directions of effect for the association with $PM_{2.5}$ and mtDNA content, only this time a hypermethylation for the mtDNA content and a hypomethylation for $PM_{2.5}$ indicating a possible downregulation of transcription by higher mtDNA content and lower $PM_{2.5}$ exposure, suggestive for the presence of potential compensatory mechanisms. As the DMR defining CpG sites were present on an island (and shore) spread over the TSS200/TSS1500 5'UTR and 1^{st} exon, a transcriptional regulation could be possible. Another overlap was found for *CCKBR*, encoding the G-protein coupled cholecystokinin B receptor which is found in the central nervous system and the gastrointestinal tract. Cancers of the colorectum, lung and pancreas have been shown to overexpress *CCKBR* (140). Here, we found a positive effect for both associations. The DMR covered the gene body and 3'UTR of *CCKBR* with the individual CpG sites located within an CpG island and shores. Because the effect of hypermethylation of CpG islands/shores within the gene body can be bi-directional, speculations about possible transcriptional outcomes are difficult.

In the present study we investigated whether mtDNA content mediates some of the effects of prenatal $PM_{2.5}$ exposure on the methylation status of CpG sites. Assuming causality, we observed that mtDNA content mediates a proportion of the effects on a CpG site located on *HNRNPM.* We are aware of the fact that a mediating effect could as easily exist in the other direction, namely for the effect of $PM_{2.5}$ exposure on mtDNA content through methylation. This has previously been investigated for the methylation of two mitochondrial DNA regions in placental tissue (50), but biological plausibility also exist for a mediation by mtDNA content. mtDNA is especially prone to damage caused by oxidative stress as explained above. The mtDNA damage as characterized by mtDNA content results in increased mitochondrial dysfunction with increased mitochondrial ROS production. Indeed, exposure to $PM_{2.5}$ has been shown to result in the generation of ROS from complex III in the mitochondrial electron transport chain (141). Besides the direct production of ROS by PM especially the secondary mitochondrial ROS can therefore affect the epigenetic landscape. The presence of high levels of ROS could result in hypermethylation of CpG sites as has been suggested (142). Various studies investigating epigenetic changes in carcinogenesis have discovered that ROS induces hypermethylation of CpG sites in the promoter regions of tumor suppressor genes and antioxidant genes resulting in gene silencing (143, 144). It has been suggested, that one possible regulatory mechanism by which ROS exerts its influence on promotor methylation could be the activation of apoptosis signaling kinase 1 (ASK1) which in turn activates JNK-dependent transcription and expression of DNMT1 resulting in gene silencing (141). Additionally, ROS increases hypermethylation by enhancing succinate and fumarate concentrations which in turn inhibit

histone demethylases (112). On the other hand, evidence also exists that PM induced mitochondrial dysfunction could result in hypomethylation. The ability of $PM_{2.5}$ to cause mitochondrial dysfunction results in a lack of ATP, which can jeopardize the cells ability to support conversion of methionine into SAM as the primary methyl donor utilized in methylation reactions (146). mtDNA content as a mediator of the effects of prenatal $PM_{2.5}$ exposure on differential methylation of CpG sites has to our knowledge not been described yet.

We acknowledge specific strengths and limitations in the present study. This is the first study to describe the association between epigenome-wide DNA methylation and *in utero* $PM_{2.5}$ exposure as well as mtDNA content in cord blood form the same population. The EWAS approach itself is agnostic and therefore allows for the discovery of new associations. Furthermore, the additional investigation not only on the level of individual CpG sites but also on a regional level by analyzing the DMRs provides additional biological significance to our findings. Moreover, by integrating gene-expression data and subsequent pathway analyses we ensured that epigenetic outcomes reliably represent biological consequences on a transcriptional level and therefore provide mechanistic insights. On the other hand, we could not find significant associations in the EWAS with mtDNA content after correcting for multiple testing and only partially for the identification of DMRs. Therefore, we based our further analyses on the 20 top ranked CpG sites and DMRs. The subsequent results are therefore also not significant and only describe trends. For the estimation of prenatal $PM_{2.5}$ exposure we applied a spatial temporal interpolation method based on the home address of the mother. Hereby we could not consider periods spent at other locations like the workplace and other time based activity patterns (exposure bias). In case of longer stays outside the living area the predicted $PM_{2.5}$ exposure could therefore be inaccurate. Although, we applied a formal mediation analyses, this method cannot prove the biological direction and therefore causality of the effect. Additionally, the outcome is only representative for one CpG site and interpretation complicated by the inconsistency in the direct and indirect effects which limits the generalizability of this finding. Strength and direction of the associations between the triangle of $PM_{2.5}$ exposure, mtDNA content and CpG methylation though differ between the ca. 450k CpG sites. One possible disadvantage of this study design could also be that the choice for linear regression models to investigate the associations between the different variables excludes possible non-linear trends. In this hypothesis generating approach we did not investigate possible sex-specific differences in the effect of $PM_{2.5}$ by testing separately for boys and girls. Future studies may also include non-linear models to account for all possible kinds of association and take into account sex-specific differences of exposure effects.

5 Conclusion

In this thesis we observed a negative association between $PM_{2.5}$ exposure during the entire pregnancy and mtDNA content in cord blood. A lower mtDNA content has been linked with adverse birth outcomes and developmental impairments later in life. High levels of $PM_{2.5}$ exposure during pregnancy could therefore be problematic for the fetus.

Likewise, the identification of six CpG sites significantly associated with prenatal $PM_{2.5}$ exposure in an EWAS demonstrates that higher prenatal $PM_{2.5}$ exposure may also act on the epigenetic level. The scope of involved genes suggests that inflammatory responses induced by $PM_{2.5}$ exposure could form the major link with adverse pregnancy and health outcomes observed in literature. A CpG site located on an island shore of *VWA1* was the most significant in the EWAS, and additionally identified by the DMRcate algorithm for the significant association with $PM_{2.5}$ exposure, corroborating previous reports in literature on the expression of vWf protein and $PM_{2.5}$ in adults. The capacity of vWf encoded by *VWA1* to recruit leucocytes, either directly or via platelets and Weibel-Palade bodies, could form a link between the epigenetic changes induced by $PM_{2.5}$ exposure and adverse health outcomes of increased inflammatory response. Higher prenatal $PM_{2.5}$ exposure may therefore contribute to adverse pregnancy outcomes and later life health problems through a transcriptional upregulation of vWf. Other genes related to inflammatory response in association with higher $PM_{2.5}$ exposure were *HNRNPM and PSG11. HNRNPM,* a mediator of inflammatory response, already found to be hypermethylated in association with cigarette smoking also showed hypermethylation in our study, implying comparable mechanisms contributing to $PM_{2.5}$ induced inflammatory processes. *PSG11,* coding a pregnancy-specific glycoprotein (PSG) which induces the secretion of anti-inflammatory cytokines via TGF-β1 during pregnancy was also found to be hypermethylated indicating an impact of higher $PM_{2.5}$ exposure levels on normal embryogenesis and innate immune regulation. As changes in the expression level of *PSG11* as well as $PM_{2.5}$ have already been associated with adverse pregnancy outcomes like low birthweight, pre-term birth and preeclampsia the differential methylation found in this study could form a link between exposure and health outcomes. Among the top 20 ranked individual CpG sites also other genes associated with the immune system like *NF115* and *UNKL* or linked to gene transcription and DNA replication such as *RING1* and *CINP*, were found to be differentially methylated, indicating additional processes linking $PM_{2.5}$ exposure with biological outcomes.

When translated to the transcriptomic level several enriched pathways comprising histone modifications, mitosis and cell cycle overlapped for the transcripts correlated with the top 20 CpG sites from EWAS of $PM_{2.5}$ exposure and mtDNA content, indicating possible intersections between these two mechanisms. Among these pathways 'HDACs deacetylate histones' and 'RMTs methylate histone arginines' promote coiling of the chromatin structures and therefore transcriptional silencing. They are dependent on intermediate metabolites provided by mitochondrial function and vice versa inhibited by increases in ROS, succinate and fumarate caused by mitochondrial dysfunction. As $PM_{2.5}$ induced oxidative stress can cause mitochondrial dysfunction with reduced production of intermediary metabolites and increased levels of ROS these mechanisms could link $PM_{2.5}$ to changes in the epigenetic landscape. Similarly, mitosis and cell cycle progression are affected by mtDNA content as well as $PM_{2.5}$. Besides previously suggested direct effects of $PM_{2.5}$ on the mitotic spindle organization resulting in cell cycle arrest and chromosomal aberrations also indirect effects through mitochondrial dysfunction and decreasing mitochondrial ATP levels form a potential link between $PM_{2.5}$ exposure and these energy dependent cellular processes. We also observed one

common gene-disease association, 'Systemic lupus erythematosus' (SLE), with an association to $PM_{2.5}$ as well as mtDNA content. SLE has previously been linked to $PM_{2.5}$, differential DNA methylation and mitochondrial dysfunction separately, however, our findings suggest that beyond this also potential links between these factors contribute to the development and etiopathology of SLE through the mechanisms of oxidative stress. Though our findings of common pathways are in line with reports in literature on the effects of $PM_{2.5}$ exposure and mtDNA content they have to be nuanced due to the fact that they are based on the outcome of CpG sites which were largely non-significant in the EWAS.

The outcome on the DMR level confirmed the results of the EWAS and pathway analysis. With regard to histone modifications differential methylation of a histone deacetylase and demethylase were found along with evidence for the hypermethylation of *TET1,* responsible for the transcription independent demethylation of DNA, linking the effects of $PM_{2.5}$ exposure with changes in global DNA methylation. A connection between $PM_{2.5}$ exposure and mtDNA content could also be established by the DMRcate algorithm in form of the most significant DMR for the association with $PM_{2.5}$. The CpG sites constituting this DMR were located on an exon of *RPTOR*, which encodes a component of the mTORC1 complex involved in the prevention of mitochondrial autophagy. The hypermethylation we found in a CpG island and shores of the gene could therefore suggest induction of increased mitochondrial autophagy by higher $PM_{2.5}$ exposure. Furthermore, we also observed an overlap of DMRs located on genes involved in cell cycle progression, innate immunity and mitochondrial energy metabolism with the bumphunter algorithm. *STIL* facilitates chromosome segregation and cell cycle progression. Hypermethylation at a CpG island upstream of *STIL* suggests a transcriptional downregulation by higher $PM_{2.5}$ exposure and mtDNA content. For *HLA-DRB6*, a highly variable pseudogene of the human major histocompatibility complex (MHC) an opposite direction of effects was observed, indicating that higher $PM_{2.5}$ exposure and lower mtDNA content induce a transcriptional downregulation. For *CPT1B,* which encodes for an enzyme facilitating the transport of long-chain fatty acyl-CoAs from cytoplasm into mitochondria, a possible downregulation of transcription by higher mtDNA content and lower $PM_{2.5}$ exposure was observed, which could be suggestive for the presence of potential compensatory mechanisms.

Finally, mediation analysis indicated that mtDNA content may mediate some of the effects of *in utero* $PM_{2.5}$ exposure on the methylation status of CpG sites. This finding is so far in line with our other outcomes as they provide evidence for the role of $PM_{2.5}$ induced mitochondrial dysfunction in the regulation of the epigenetic landscape but interpretations should be made in the light of the limitations of the analysis regarding translational significance, as the analysis was restricted only to one CpG site.

In this study we identified several CpG sites and DMRs significantly associated with prenatal $PM_{2.5}$ exposure and found evidence for a significant effect of $PM_{2.5}$ on cord blood mtDNA content. Our findings further indicate a mediation of effects of $PM_{2.5}$ on DNA methylation by mtDNA content. This study also suggests mechanisms acting on methylation through mitochondrial energy supply, mitotic progression and innate immunity. So far, we achieved our initial goal and corroborated the hypothesis that we can identify CpGs and differentially methylated regions (DMRs) associated with prenatal $PM_{2.5}$ exposure and mtDNA content. Future studies are needed to further investigate the potential role of the discovered candidate genes and pathways and the relevance of mitochondrial (dys)function in the occurrence of epigenetic changes and adverse health outcomes in response to *in utero* PM2.5 exposure.

6 References

1. Lewtas J. Air pollution combustion emissions: Characterization of causative agents and mechanisms associated with cancer, reproductive, and cardiovascular effects. Mutation Research/Reviews in Mutation Research. 2007;636(1):95-133.
2. Valavanidis A, Fiotakis K, Vlachogianni T. Airborne particulate matter and human health: toxicological assessment and importance of size and composition of particles for oxidative damage and carcinogenic mechanisms. Journal of environmental science and health Part C, Environmental carcinogenesis & ecotoxicology reviews. 2008;26(4):339-62.
3. Xing Y-F, Xu Y-H, Shi M-H. The impact of PM2.5 on the human respiratory system. Journal of Thoracic Disease. 2016;8(1):E69-E74.
4. Geiser M, Rothen-Rutishauser B, Kapp N, Schürch S, Kreyling W, Schulz H, et al. Ultrafine Particles Cross Cellular Membranes by Nonphagocytic Mechanisms in Lungs and in Cultured Cells. Environmental health perspectives. 2005;113(11):1555-60.
5. World Health Organization. Department of Public Health E, Health SDo. Ambient Air Pollution: A Global Assessment of Exposure and Burden of Disease: World Health Organization; 2016.
6. Shah ASV, Lee KK, McAllister DA, Hunter A, Nair H, Whiteley W, et al. Short term exposure to air pollution and stroke: systematic review and meta-analysis. BMJ (Clinical research ed). 2015;350.
7. Dockery DW, Pope CA, 3rd. Acute respiratory effects of particulate air pollution. Annual review of public health. 1994;15:107-32.
8. Scheers H, Jacobs L, Casas L, Nemery B, Nawrot TS. Long-Term Exposure to Particulate Matter Air Pollution Is a Risk Factor for Stroke: Meta-Analytical Evidence. Stroke. 2015;46(11):3058-66.
9. Pope CA, 3rd, Burnett RT, Thun MJ, Calle EE, Krewski D, Ito K, et al. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. Jama. 2002;287(9):1132-41.
10. Brook RD, Rajagopalan S, Pope CA, Brook JR, Bhatnagar A, Diez-Roux AV, et al. Particulate Matter Air Pollution and Cardiovascular Disease. An Update to the Scientific Statement From the American Heart Association. 2010;121(21):2331-78.
11. Raaschou-Nielsen O, Andersen ZJ, Beelen R, et al. Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE). The Lancet Oncology. 2013;14(9):813-22.
12. Pope IC, Burnett RT, Thun MJ, et al. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. Jama. 2002;287(9):1132-41.
13. Loomis D, Grosse Y, Lauby-Secretan B, El Ghissassi F, Bouvard V, Benbrahim-Tallaa L, et al. The carcinogenicity of outdoor air pollution. The Lancet Oncology. 2013;14(13):1262-3.
14. Qian Z, Liang S, Yang S, Trevathan E, Huang Z, Yang R, et al. Ambient air pollution and preterm birth: A prospective birth cohort study in Wuhan, China. International Journal of Hygiene and Environmental Health. 2016;219(2):195-203.
15. Rappazzo KM, Daniels JL, Messer LC, Poole C, Lobdell DT. Exposure to fine particulate matter during pregnancy and risk of preterm birth among women in New Jersey, Ohio, and Pennsylvania, 2000-2005. Environmental health perspectives. 2014;122(9):992-7.
16. Li X, Huang S, Jiao A, Yang X, Yun J, Wang Y, et al. Association between ambient fine particulate matter and preterm birth or term low birth weight: An updated systematic review and meta-analysis. Environmental Pollution. 2017;227(Supplement C):596-605.
17. Kingsley SL, Eliot MN, Glazer K, Awad YA, Schwartz JD, Savitz DA, et al. Maternal ambient air pollution, preterm birth and markers of fetal growth in Rhode Island: results of a hospital-based linkage study. Journal of epidemiology and community health. 2017.
18. Barker DJ. Fetal origins of coronary heart disease. BMJ (Clinical research ed). 1995;311(6998):171-4.
19. Heck JE, Wu J, Lombardi C, Qiu J, Meyers TJ, Wilhelm M, et al. Childhood Cancer and Traffic-Related Air Pollution Exposure in Pregnancy and Early Life. Environmental health perspectives. 2013;121(11-12):1385-91.
20. Lavigne E, Belair MA, Do MT, Stieb DM, Hystad P, van Donkelaar A, et al. Maternal exposure to ambient air pollution and risk of early childhood cancers: A population-based study in Ontario, Canada. Environment international. 2017;100:139-47.
21. Suades-Gonzalez E, Gascon M, Guxens M, Sunyer J. Air Pollution and Neuropsychological Development: A Review of the Latest Evidence. Endocrinology. 2015;156(10):3473-82.
22. Hsu HH, Chiu YH, Coull BA, Kloog I, Schwartz J, Lee A, et al. Prenatal Particulate Air Pollution and Asthma Onset in Urban Children. Identifying Sensitive Windows and Sex Differences. American journal of respiratory and critical care medicine. 2015;192(9):1052-9.
23. Jedrychowski WA, Perera FP, Maugeri U, Mroz E, Klimaszewska-Rembiasz M, Flak E, et al. EFFECT OF PRENATAL EXPOSURE TO FINE PARTICULATE MATTER ON VENTILATORY LUNG FUNCTION OF PRESCHOOL CHILDREN OF NONSMOKING MOTHERS. KRAKOW INNER CITY BIRTH COHORT PROSPECTIVE STUDY. Paediatric and perinatal epidemiology. 2010;24(5):492-501.
24. Jedrychowski WA, Perera FP, Spengler JD, Mroz E, Stigter L, Flak E, et al. Intrauterine exposure to fine particulate matter as a risk factor for increased susceptibility to acute broncho-pulmonary infections in early childhood. Int J Hyg Environ Health. 2013;216(4):395-401.
25. Tanwar V, Gorr MW, Velten M, Eichenseer CM, Long VP, 3rd, Bonilla IM, et al. In Utero Particulate Matter Exposure Produces Heart Failure, Electrical Remodeling, and Epigenetic Changes at Adulthood. Journal of the American Heart Association. 2017;6(4).
26. Wick P, Malek A, Manser P, Meili D, Maeder-Althaus X, Diener L, et al. Barrier Capacity of Human Placenta for Nanosized Materials. Environmental health perspectives. 2010;118(3):432-6.
27. Zhang X, Li X, Jing Y, et al. Transplacental transfer of polycyclic aromatic hydrocarbons in paired samples of maternal serum, umbilical cord serum, and placenta in Shanghai, China. Environmental pollution (Barking, Essex : 1987). 2017;222:267-75.
28. Autrup H, Vestergaard AB, Okkels H. Transplacental transfer of environmental genotoxins: polycyclic aromatic hydrocarbon-albumin in non-smoking women, and the effect of maternal GSTM1 genotype. Carcinogenesis. 1995;16(6):1305-9.
29. Zaretsky MV, Alexander JM, Byrd W, Bawdon RE. Transfer of inflammatory cytokines across the placenta. Obstetrics and gynecology. 2004;103(3):546-50.
30. Roubicek DA, Souza-Pinto N. Mitochondria and mitochondrial DNA as relevant targets for environmental contaminants.Toxicology. 2017.
31. Alexeyev M, Shokolenko I, Wilson G, LeDoux S. The Maintenance of Mitochondrial DNA Integrity—Critical Analysis and Update. Cold Spring Harbor Perspectives in Biology. 2013;5(5):a012641.
32. Boland ML, Chourasia AH, Macleod KF. Mitochondrial dysfunction in cancer. Frontiers in oncology. 2013;3:292.
33. Lowell BB, Shulman GI. Mitochondrial dysfunction and type 2 diabetes. Science (New York, NY). 2005;307(5708):384-7.
34. Moreira PI, Carvalho C, Zhu X, Smith MA, Perry G. Mitochondrial dysfunction is a trigger of Alzheimer's disease pathophysiology. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease. 2010;1802(1):2-10.
35. Abou-Sleiman PM, Muqit MM, Wood NW. Expanding insights of mitochondrial dysfunction in Parkinson's disease. Nature reviews Neuroscience. 2006;7(3):207-19.
36. Quintanilla RA, Johnson GVW. Role of Mitochondrial Dysfunction in the Pathogenesis of Huntington's Disease. Brain research bulletin. 2009;80(4-5):242-7.
37. Figueira TR, Barros MH, Camargo AA, Castilho RF, Ferreira JC, Kowaltowski AJ, et al. Mitochondria as a source of reactive oxygen and nitrogen species: from molecular mechanisms to human health. Antioxid Redox Signal. 2013;18(16):2029-74.
38. Yakes FM, Van Houten B. Mitochondrial DNA damage is more extensive and persists longer than nuclear DNA damage in human cells following oxidative stress. Proceedings of the National Academy of Sciences of the United States of America. 1997;94(2):514-9.
39. Risom L, Moller P, Loft S. Oxidative stress-induced DNA damage by particulate air pollution. Mutation research. 2005;592(1-2):119-37.
40. Valavanidis A, Vlachogianni T. 8-hydroxy-2' -deoxyguanosine (8-OHdG): A critical biomarker of oxidative stress and carcinogenesis. Journal of environmental science and health Part C, J Environ Sci Health C Environ Carcinog Ecotoxicol Rev. 2009;27(2):120-39.
41. Cooke MS, Evans MD, Dizdaroglu M, Lunec J. Oxidative DNA damage: mechanisms, mutation, and disease. FASEB journal : official publication of the Federation of American Societies for Experimental Biology. 2003;17(10):1195-214.
42. Moller P, Danielsen PH, Karottki DG, Jantzen K, Roursgaard M, Klingberg H, et al. Oxidative stress and inflammation generated DNA damage by exposure to air pollution particles. Mutation research Reviews in mutation research. 2014;762:133-66.
43. Herbstman JB, Tang D, Zhu D, Qu L, Sjodin A, Li Z, et al. Prenatal exposure to polycyclic aromatic hydrocarbons, benzo[a]pyrene-DNA adducts, and genomic DNA methylation in cord blood. Environmental health perspectives. 2012;120(5):733-8.
44. Sram RJ, Binkova B, Beskid O, Milcova A, Rossner P, Rossner P, Jr., et al. Biomarkers of exposure and effect-interpretation in human risk assessment. Air quality, atmosphere, & health. 2011;4(3-4):161-7.
45. Grevendonk L, Janssen BG, Vanpoucke C, Lefebvre W, Hoxha M, Bollati V, et al. Mitochondrial oxidative DNA damage and exposure to particulate air pollution in mother-newborn pairs. Environmental Health. 2016;15:10.
46. Lim KS, Jeyaseelan K, Whiteman M, Jenner A, Halliwell B. Oxidative damage in mitochondrial DNA is not extensive. Annals of the New York Academy of Sciences. 2005;1042:210-20.
47. Lee H-C, Wei Y-H. Mitochondrial biogenesis and mitochondrial DNA maintenance of mammalian cells under oxidative stress. The International Journal of Biochemistry & Cell Biology. 2005;37(4):822-34.

48. Lee J-W, Park KD, Im J-A, Kim MY, Lee D-C. Mitochondrial DNA copy number in peripheral blood is associated with cognitive function in apparently healthy elderly women. Clinica Chimica Acta. 2010;411(7):592-6.
49. Malik AN, Czajka A.Is mitochondrial DNA content a potential biomarker of mitochondrial dysfunction?Mitochondrion.2013;13(5):481-92.
50. Janssen BG, Byun HM, Gyselaers W, Lefebvre W, Baccarelli AA, Nawrot TS. Placental mitochondrial methylation and exposure to airborne particulate matter in the early life environment: An ENVIRONAGE birth cohort study. Epigenetics. 2015;10(6):536-44.
51. Rosa MJ, Just AC, Guerra MS, Kloog I, Hsu HL, Brennan KJ, et al. Identifying sensitive windows for prenatal particulate air pollution exposure and mitochondrial DNA content in cord blood. Environment international. 2017;98:198-203.
52. Hales BF, Grenier L, Lalancette C, Robaire B. Epigenetic programming: from gametes to blastocyst. Birth defects research Part A, Clinical and molecular teratology. 2011;91(8):652-65.
53. Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat Genet. 2003;33 Suppl:245
54. Shock LS, Thakkar PV, Peterson EJ, Moran RG, Taylor SM. DNA methyltransferase 1, cytosine methylation, and cytosine hydroxymethylation in mammalian mitochondria. Proceedings of the National Academy of Sciences of the United States of America. 2011;108(9):3630-5.
55. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science (New York, NY). 2009;324(5929):930-5.
56. Valinluck V, Tsai HH, Rogstad DK, Burdzy A, Bird A, Sowers LC. Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2). Nucleic acids research. 2004;32(14):4100-8.
57. Baccarelli A, Bollati V. Epigenetics and environmental chemicals. Current opinion in pediatrics. 2009;21(2):243-51.
58. Dai L, Mehta A, Mordukhovich I, Just AC, Shen J, Hou L, et al. Differential DNA methylation and PM2.5 species in a 450K epigenome-wide association study. Epigenetics. 2017;12(2):139-48.
59. Plusquin M, Guida F, Polidoro S, Vermeulen R, Raaschou-Nielsen O, Campanella G, et al. DNA methylation and exposure to ambient air pollution in two prospective cohorts. Environment international. 2017;108:127-36.
60. Janssen BG, Godderis L, Pieters N, Poels K, Kicinski M, Cuypers A, et al. Placental DNA hypomethylation in association with particulate air pollution in early life. Particle and fibre toxicology. 2013;10:22.
61. Bhasin JM, Hu B, Ting AH. MethylAction: detecting differentially methylated regions that distinguish biological subtypes. Nucleic Acids Research. 2016;44(1):106-16.
62. Janssen BG, Madlhoum N, Gyselaers W, Bijnens E, Clemente DB, Cox B, et al. Cohort Profile: The ENVIRonmental influence ON early AGEing (ENVIRONAGE): a birth cohort study. International journal of epidemiology. 2017.
63. Vineis P, Chadeau-Hyam M, Gmuender H, Gulliver J, Herceg Z, Kleinjans J, et al. The exposome in practice: Design of the EXPOsOMICS project. International Journal of Hygiene and Environmental Health. 2017;220(2):142-51.
64.World Medical Association Declaration of Helsinki:ethical principles for medical research involving human subjects.Jama.2013;310(20):2191-4.
65. Janssen S, Dumont G, Fierens F, Mensink C. Spatial interpolation of air pollution measurements using CORINE land cover data. Atmospheric Environment. 2008;42(20):4884-903.
66. Bibikova M, Barnes B, Tsan C, et al. High density DNA methylation array with single CpG site resolution. Genomics. 2011;98(4):288-95.
67. Bakulski KM, Feinberg JI, Andrews SV, Yang J, Brown S, L. McKenney S, et al. DNA methylation of cord blood cell types: Applications for mixed cell birth studies. Epigenetics. 2016;11(5):354-62.
68. Chadeau-Hyam M, Vermeulen RC, Hebels DG, et al. Prediagnostic transcriptomic markers of Chronic lymphocytic leukemia reveal perturbations 10 years before diagnosis. Annals of oncology: official journal of the European Society for Medical Oncology. 2014;25(5):1065-72.
69. Huang DW, Sherman BT, Tan Q, Collins JR, Alvord WG, Roayaei J, et al. The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. Genome biology. 2007;8(9):R183-R.
70. Peters TJ, Buckley MJ, Statham AL, Pidsley R, Samaras K, V Lord R, et al. De novo identification of differentially methylated regions in the human genome. Epigenetics & Chromatin. 2015;8(1):6.
71. Jaffe AE, Murakami P, Lee H, Leek JT, Fallin MD, Feinberg AP, et al. Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. International journal of epidemiology. 2012;41(1):200-9.
72. Du P, Zhang X, Huang C-C, Jafari N, Kibbe WA, Hou L, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformatics. 2010;11:587-.
73. Chen Y-a, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics. 2013;8(2):203-9.
74. Akulenko R, Merl M, Helms V. BEclear: Batch Effect Detection and Adjustment in DNA Methylation Data. PloS one. 2016;11(8):e0159921.
75. Infinium HumanMethylation450K v1.2 Product Files 2014 [Available from:http://support.illumina.com/downloads/infinium_humanmethylation450_product_files.html.
76. MacKinnon D. Introduction to statistical mediation analysis: Routledge; 2012.
77. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. Journal of personality and social psychology. 1986;51(6):1173-82.
78. Pearl J. Direct and indirect effects. Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence; Seattle, Washington. 2074073: Morgan Kaufmann Publishers Inc.; 2001. p. 411-20.
79. Valeri L, Vanderweele TJ. Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. Psychological methods. 2013;18(2):137-50.
80. MacKinnon DP, Krull JL, Lockwood CM. Equivalence of the Mediation, Confounding and Suppression Effect. Prevention science : the official journal of the Society for Prevention Research. 2000;1(4):173.
81. Alwin DF, Hauser RM. The Decomposition of Effects in Path Analysis. American Sociological Review. 1975;40(1):37-47.
82. Pieters N, Janssen BG, Dewitte H, Cox B, Cuypers A, Lefebvre W, et al. Biomolecular Markers within the Core Axis of Aging and Particulate Air Pollution Exposure in the Elderly: A Cross-Sectional Study. Environmental health perspectives. 2016;124(7):943-50.
83. Clemente DBP, Casas M, Vilahur N, Begiristain H, et al. Prenatal Ambient Air Pollution, Placental Mitochondrial DNA Content, and Birth Weight in the INMA (Spain) and ENVIRONAGE (Belgium) Birth Cohorts. Environmental health perspectives. 2016;124(5):659-65.
84. Clemente DBP, Casas M, Janssen BG, Lertxundi A, Santa-Marina L, Iñiguez C, et al. Prenatal ambient air pollution exposure, infant growth and placental mitochondrial DNA content in the INMA birth cohort. Environmental research. 2017;157:96-102.
85. F. GT, Silvia S, Guillermo D, I. GS, Carolina G, D. GC, et al. A Decreased Mitochondrial DNA Content Is Related to Insulin Resistance in Adolescents. Obesity. 2008;16(7):1591-5.
86. Panni T, Mehta AJ, Schwartz JD, Baccarelli AA, Just AC, Wolf K, et al. Genome-Wide Analysis of DNA Methylation and Fine Particulate Matter Air Pollution in Three Study Populations:KORA F3, KORA F4, and the Normative Aging Study. Environmental health perspectives. 2016;124(7):983-90.
87. Pontarotti P. Evolutionary Biology: Convergent Evolution, Evolution of Complex Traits, Concepts and Methods: Springer International Publishing; 2016.
88. Pihl K, Larsen T, Laursen I, Krebs L, Christiansen M. First trimester maternal serum pregnancy-specific beta-1-glycoprotein (SP1) as a marker of adverse pregnancy outcome. Prenatal diagnosis. 2009;29(13):1256-61.
89. Zhao L, Triche EW, Walsh KM, Bracken MB, Saftlas AF, Hoh J, et al. Genome-wide association study identifies a maternal copy-number deletion in PSG11 enriched among preeclampsia patients. BMC Pregnancy Childbirth. 2012;12:61.
90. Wu J, Ren C, Delfino RJ, Chung J, Wilhelm M, Ritz B. Association between Local Traffic-Generated Air Pollution and Preeclampsia and Preterm Delivery in the South Coast Air Basin of California. Environmental health perspectives. 2009;117(11):1773-9.
91. Dadvand P, Figueras F, Basagaña X, Beelen R, Martinez D, Cirach M, et al. Ambient Air Pollution and Preeclampsia: A Spatiotemporal Analysis. Environmental health perspectives. 2013;121(11-12):1365-71.
92. Thomas P, Forse RA, Bajenova O. Carcinoembryonic antigen (CEA) and its receptor hnRNP M are mediators of metastasis and the inflammatory response in the liver. Clinical & experimental metastasis. 2011;28(8):923-32.
93. Buro-Auriemma LJ, Salit J, Hackett NR, Walters MS, Strulovici-Barel Y, Staudt MR, et al. Cigarette smoking induces small airway epithelial epigenetic changes with corresponding modulation of gene expression. Human Molecular Genetics. 2013;22(23):4726-38.
94. Sadler JE. Biochemistry and Genetics of von Willebrand Factor. Annual review of biochemistry. 1998;67(1):395-424.
95. Conway DS, Pearce LA, Chin BS, Hart RG, Lip GY. Prognostic value of plasma von Willebrand factor and soluble P-selectin as indices of endothelial damage and platelet activation in 994 patients with nonvalvular atrial fibrillation. Circulation. 2003;107(25):3141-5.
96. Ruckerl R, Ibald-Mulli A, Koenig W, Schneider A, Woelke G, Cyrys J, et al. Air pollution and markers of inflammation and coagulation in patients with coronary heart disease. American journal of respiratory and critical care medicine. 2006;173(4):432-41.
97. Riediker M. Cardiovascular effects of fine particulate matter components in highway patrol officers.Inhalation toxicology 2007;19 Suppl 1:99-105.

98. Liao D, Heiss G, Chinchilli VM, Duan Y, Folsom AR, Lin HM, et al. Association of criteria pollutants with plasma hemostatic/inflammatory markers: a population-based study. Journal of exposure analysis and environmental epidemiology. 2005;15(4):319-28.
99. Rich DQ, Kipen HM, Huang W, et al. Association between changes in air pollution levels during the beijing olympics and biomarkers of inflammation and thrombosis in healthy young adults. Jama. 2012;307(19):2068-78.
100. Yuan Z, Chen Y, Zhang Y, Liu H, Liu Q, Zhao J, et al. Changes of plasma vWF level in response to the improvement of air quality: an observation of 114 healthy young adults. Annals of hematology. 2013;92(4):543-8.
101. Rao X, Evans J, Chae H, Pilrose J, Kim S, Yan P, et al. CpG island shore methylation regulates caveolin-1 expression in breast cancer. Oncogene. 2013;32(38):4519-28.
102. Martino D, Saffery R. Characteristics of DNA methylation and gene expression in regulatory features on the Infinium 450k Beadchip. bioRxiv. 2015.
103. Jones PA. The DNA methylation paradox. Trends in genetics : TIG. 1999;15(1):34-7.
104. Saenen ND, Vrijens K, Janssen BG, Roels HA, Neven KY, Vanden Berghe W, et al. Lower Placental Leptin Promoter Methylation in Association with Fine Particulate Matter Air Pollution during Pregnancy and Placental Nitrosative Stress at Birth in the ENVIRONAGE Cohort. Environmental health perspectives. 2017;125(2):262-8.
105. Li F, Wang Y, Zeller KI, Potter JJ, Wonsey DR, O'Donnell KA, et al. Myc Stimulates Nuclearly Encoded Mitochondrial Genes and Mitochondrial Biogenesis. Molecular and Cellular Biology. 2005;25(14):6225-34.
106. Morrish F, Hockenbery D. MYC and Mitochondrial Biogenesis. Cold Spring Harbor Perspectives in Medicine. 2014;4(5):a014225.
107. Scarpulla RC. Metabolic control of mitochondrial biogenesis through the PGC-1 family regulatory network. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research. 2011;1813(7):1269-78.
108.Tessarz P, Kouzarides T. Histone core modifications regulating nucleosome structure and dynamics.Nat Rev Mol Cell Biol.2014;15(11):703-8.
109. Cao D, Bromberg PA, Samet JM. COX-2 expression induced by diesel particles involves chromatin modification and degradation of HDAC1. American journal of respiratory cell and molecular biology. 2007;37(2):232-9.
110. Adenuga D, Yao H, March TH, Seagrave J, Rahman I. Histone Deacetylase 2 Is Phosphorylated, Ubiquitinated, and Degraded by Cigarette Smoke. American journal of respiratory cell and molecular biology. 2009;40(4):464-73.
111. Martinez-Reyes I, Diebold LP, Kong H, Schieber M, Huang H, Hensley CT, et al. TCA Cycle and Mitochondrial Membrane Potential Are Necessary for Diverse Biological Functions. Molecular cell. 2016;61(2):199-209.
112. Matilainen O, Quiros PM, Auwerx J .Mitochondria and Epigenetics-Crosstalk in Homeostasis and Stress.Trends Cell Biol.2017;27(6):453-63.
113. Gualtieri M, Ovrevik J, Mollerup S, Asare N, Longhin E, Dahlman HJ, et al. Airborne urban particles (Milan winter-PM2.5) cause mitotic arrest and cell death: Effects on DNA, mitochondria, AhR binding and spindle organization. Mutation research. 2011;713(1-2):18-31.
114. Longhin E, Holme JA, Gutzkow KB, Arlt VM, Kucab JE, Camatini M, et al. Cell cycle alterations induced by urban PM2.5 in bronchial epithelial cells: characterization of the process and possible mechanisms involved. Particle and fibre toxicology. 2013;10:63-.
115. Wang Z, Fan M, Candas D, Zhang TQ, Qin L, Eldridge A, et al. Cyclin B1/Cdk1 coordinates mitochondrial respiration for cell-cycle G2/M progression. Developmental cell. 2014;29(2):217-32.
116. Richardson B, Scheinbart L, Strahler J, Gross L, Hanash S, Johnson M. Evidence for impaired T cell DNA methylation in systemic lupus erythematosus and rheumatoid arthritis. Arthritis Rheum. 1990;33(11):1665-73.
117. Balada E, Ordi-Ros J, Vilardell-Tarres M. DNA methylation and systemic lupus erythematosus. Annals of the New York Academy of Sciences. 2007;1108:127-36.
118. Sekigawa I, Kawasaki M, Ogasawara H, Kaneda K, Kaneko H, Takasaki Y, et al. DNA methylation: its contribution to systemic lupus erythematosus. Clinical and experimental medicine. 2006;6(3):99-106.
119. Fernandes EC, Silva CA, Braga AL, Sallum AM, Campos LM, Farhat SC. Exposure to Air Pollutants and Disease Activity in Juvenile-Onset Systemic Lupus Erythematosus Patients. Arthritis care & research. 2015;67(11):1609-14.
120. Alves AGF, de Azevedo Giacomin MF, Braga ALF, Sallum AME, Pereira LAA, Farhat LC, et al. Influence of air pollution on airway inflammation and disease activity in childhood-systemic lupus erythematosus. Clinical rheumatology. 2018;37(3):683-90.
121. Gergely P, Grossman C, Niland B, Puskas F, Neupane H, Allam F, et al. Mitochondrial Hyperpolarization and ATP Depletion in Patients With Systemic Lupus Erythematosus. Arthritis and rheumatism. 2002;46(1):175-90.
122. Perl A, Hanczko R, Doherty E. Assessment of mitochondrial dysfunction in lymphocytes of patients with systemic lupus erythematosus. Methods in molecular biology (Clifton, NJ). 2012;900:61-89.
123. Ben-Chetrit E, Chan EK, Sullivan KF, A TE. A 52-kD protein is a novel component of the SS-A/Ro antigenic particle. The Journal of Experimental Medicine. 1988;167(5):1560-71.
124. Jaffe AE, Feinberg AP, Irizarry RA, Leek JT. Significance analysis and statistical dissection of variably methylated regions. Biostatistics (Oxford, England). 2012;13(1):166-78.
125. Morita M, Gravel S-P, Chénard V, Sikström K, Zheng L, Alain T, et al. mTORC1 Controls Mitochondrial Activity and Biogenesis through 4E-BP-Dependent Translational Regulation. Cell Metabolism. 2013;18(5):698-711.
126. Wullschleger S, Loewith R, Hall MN. TOR signaling in growth and metabolism. Cell. 2006;124(3):471-84.
127. Polak P, Cybulski N, Feige JN, Auwerx J, Rüegg MA, Hall MN. Adipose-Specific Knockout of raptor Results in Lean Mice with Enhanced Mitochondrial Respiration. Cell Metabolism. 2008;8(5):399-410.
128. Gruzieva O, Xu CJ, Breton CV, Annesi-Maesano I, Anto JM, Auffray C, et al. Epigenome-Wide Meta-Analysis of Methylation in Children Related to Prenatal NO2 Air Pollution Exposure. Environmental health perspectives. 2017;125(1):104-10.
129. Unoki M, Nakamura Y. Methylation at CpG islands in intron 1 of EGR2 confers enhancer-like activity. FEBS letters. 2003;554(1):67-72.
130. Li X, Liu L, Yang S, Song N, Zhou X, Gao J, et al. Histone demethylase KDM5B is a key regulator of genome stability. Proceedings of the National Academy of Sciences. 2014;111(19):7096-101.
131. Schübeler D. Function and information content of DNA methylation. Nature. 2015;517:321.
132. Li H, Wen WP. Fine particulate air pollution epigenetically affects TET1 expression in dendritic cells and exacerbates the inflammation of Allergic Rhinitis. Journal of Allergy and Clinical Immunology. 2018;141(2, Supplement):AB75.
133. Sanchez-Guerra M, Zheng Y, Osorio-Yanez C, Zhong J, Chervona Y, Wang S, et al. Effects of particulate matter exposure on blood 5-hydroxymethylation: results from the Beijing truck driver air pollution study. Epigenetics. 2015;10(7):633-42.
134. Rahman I, Marwick J, Kirkham P. Redox modulation of chromatin remodeling: impact on histone acetylation and deacetylation, NF-kappaB and pro-inflammatory gene expression. Biochemical pharmacology. 2004;68(6):1255-67.
135.Licciardi PV, Karagiannis TC. Regulation of Immune Responses by Histone Deacetylase Inhibitors. ISRN Hematology. 2012;2012:690901.
136. Kumar A, Girimaji SC, Duvvari MR, Blanton SH. Mutations in STIL, Encoding a Pericentriolar and Centrosomal Protein, Cause Primary Microcephaly. American journal of human genetics. 2009;84(2):286-90.
137. Daude N, Wohlgemuth S, Brown R, Pitstick R, Gapeshina H, Yang J, et al. Knockout of the prion protein (PrP)-like Sprn gene does not produce embryonic lethality in combination with PrP(C)-deficiency. Proceedings of the National Academy of Sciences of the United States of America. 2012;109(23):9035-40.
138. Fernandez-Soria VM, Morales P, Castro MJ, Suarez B, Recio MJ, Moreno MA, et al. Transcription and weak expression of HLA-DRB6: a gene with anomalies in exon 1 and other regions. Immunogenetics. 1998;48(1):16-21.
139. van der Leij FR, Cox KB, Jackson VN, Huijkman NCA, Bartelds B, Kuipers JRG, et al. Structural and Functional Genomics of the CPT1B Gene for Muscle-type Carnitine Palmitoyltransferase I in Mammals. Journal of Biological Chemistry. 2002;277(30):26994-7005.
140. Roy J, Putt KS, Coppola D, Leon ME, Khalil FK, Centeno BA, et al. Assessment of cholecystokinin 2 receptor (CCK2R) in neoplastic tissue. Oncotarget. 2016;7(12):14605-15.
141. Soberanes S, Urich D, Baker CM, et al. Mitochondrial Complex III-generated Oxidants Activate ASK1 and JNK to Induce Alveolar Epithelial Cell Death following Exposure to Particulate Matter Air Pollution. The Journal of Biological Chemistry. 2009;284(4):2176-86.
142. Smiraglia DJ, Kulawiec M, Bistulfi GL, Gupta SG, Singh KK. A novel role for mitochondria in regulating epigenetic modification in the nucleus. Cancer biology & therapy. 2008;7(8):1182-90.
143. Franco R, Schoneveld O, Georgakilas AG. Oxidative stress, DNA methylation and carcinogenesis. Cancer letters. 2008;266(1):6-11.
144. Ziech D, Franco R, Pappa A, Panayiotidis MI. Reactive oxygen species (ROS)--induced genetic and epigenetic alterations in human carcinogenesis. Mutation research. 2011;711(1-2):167-73.
145. Soberanes S, Gonzalez A, Urich D, Chiarella SE, Radigan KA, Osornio-Vargas A, et al. Particulate matter Air Pollution induces hypermethylation of the p16 promoter Via a mitochondrial ROS-JNK-DNMT1 pathway. Scientific Reports. 2012;2:275.
146. Minocherhomji S,Tollefsbol TO, Singh KK.Mitochondrial regulation of epigenetics and its role in human diseases. Epigenetics. 2012;7(4)

SUPPLEMENTAL MATERIAL

Epigenome-wide DNA methylation and mitochondrial DNA content in relation to *in utero* $PM_{2.5}$ exposure

Senior Practical Training
Master Biomedical Sciences
Environmental Health Sciences
November 2017 – June 2018

Centre For Environmental Sciences (CMK)
Supervisor: Prof. dr. Michelle Plusquin
Co-Supervisor: Dr. Bram Janssen
Daily Supervisor: Annette Vriens

Brigitte Reimann (1334172)

Table of Contents

Supplementary tables

Table S1. Primer sequences for mitochondrial and nuclear genes.

Gene	Forward 5'–3'	Reverse 5'–3'	Primer efficiency (%)
MTF3212/R3319	CACCCAAGAACAGGGTTTGT	TGGCCATGGGTATGTTGTTAA	96.3
MT-ND1	ATGGCCAACCTCCTACTCCT	CTACAACGTTGGGGCCTTT	99.3
RPLP0	GGAATGTGGGCTTTGTGTTC	CCCAATTGTCCCCTTACCTT	100.7
ACTB	ACTCTTCCAGCCTTCCTTCC	GGCAGGACTTAGCTTCCACA	96.8

Abbreviations: ACTB, beta actin; MTF3212/R3319, mitochondrial forward primer from nucleotide 3212 and reverse primer from nucleotide 3319; MT-ND1, mitochondrial encoded NADH dehydrogenase 1; RPLP0, acidic ribosomal phosphoprotein P0.

Table S2.

Top 20 individual CpG sites highest ranked according to *p*-value for the association with $PM_{2.5}$ in model 1 of the EWAS.

CpG	Gene	Position	Localizationion on CGI	Localization on gene	ß	SE	p-Value	q-Value
cg09843049	*VWA1*	chr1: 1373316	North Shore	Body, 3'UTR	0.00568	0.00111	8.57E-07	1.83E-01
cg07452728	*PIN1*	chr19: 9960079	Open Sea	3'UTR	0.00133	0.00026	9.80E-07	1.83E-01
cg26140182	*PSG11*	chr19: 43530536	Open Sea	5'UTR, 1stExon	0.00424	0.00085	1.43E-06	1.83E-01
cg26834192	*AGPAT4*	chr6: 161561031	Open Sea	Body	0.00315	0.00064	2.25E-06	1.83E-01
cg03804903	*HNRNPM*	chr19: 8551040	Island	Body	0.0024	0.00049	2.25E-06	1.83E-01
cg05600804	—	chrX: 153623245	North Shelf		0.00304	0.00062	2.26E-06	1.83E-01
cg14173815	*RPL23A, SNORD42A*	chr17: 27049557	North Shelf	Body, TSS1500, TSS200	0.00362	0.00075	3.20E-06	2.05E-01
cg00308185	*BRDT*	chr1: 92414910	Island	TSS200	0.0033	0.00069	3.38E-06	2.05E-01
cg05534865	—	chr16: 1444162	South Shelf		0.00328	0.00069	4.34E-06	2.34E-01
cg02930037	*RPN2, C20orf132*	chr20: 35807377	North Shore	TSS200, Body	-0.00194	0.00042	7.39E-06	2.95E-01
cg08316775	*TPRKB*	chr2: 73964643	Island	TSS200	-0.00105	0.00023	7.44E-06	2.95E-01
cg16036471	—	chr5: 7271612	South Shore		-0.00304	0.00066	7.54E-06	2.95E-01
cg17448109	*RNF115*	chr1: 145670590	Open Sea	Body	0.00308	0.00067	8.45E-06	2.95E-01
cg27327785	*RING1*	chr6: 33176447	South Shore	5'UTR	-0.0018	0.00039	8.50E-06	2.95E-01
cg14554618	*NAIF1, SLC25A25*	chr9: 130829035	North Shore	1stExon, TSS1500	0.00236	0.00052	1.12E-05	3.26E-01
cg04283429	—	chr6 :99280108	Island		-0.00254	0.00056	1.14E-05	3.26E-01
cg09842285	*UNKL*	chr16: 1429729	Island	TSS200	-0.00089	0.0002	1.26E-05	3.40E-01
cg03054491	*UTP20*	chr12: 101674203	South Shore	Body	-0.00129	0.00029	1.53E-05	3.73E-01
cg04717240	*CINP, TECPR2*	chr14: 102829258	Island	TSS200, TSS200	-0.00189	0.00043	1.71E-05	3.73E-01
cg16892812	—	chr7: 121796366	Open Sea	1stExon, TSS1500	-0.00377	0.00085	1.75E-05	0.37267

The number of subjects included in the two models is n = 187. β value (0–1 scale) represents the difference in methylation for every unit ($\mu g/m^3$) increase of $PM_{2.5}$. Column headers: Gene = UCSC annotated gene; Position = chromosome and chromosomal position; Localization on CGI= UCSC gene region feature category; Localization on CGI = UCSC relation to CpG islands; β = regression coefficient; SE = standard error for regression coefficient.

Table S3.

Top 20 individual CpG sites highest ranked according to *p*-value for the association with mtDNA content in model 1 of the EWAS.

CpG	Gene	Position	Localization on CGI	Localization on gene	ß	SE	p-Value	q-Value
cg27597473	*NOTCH4*	chr6: 32172013	Open Sea	Body	-0.0278	0.0056	1.60E-06	0.465
cg24847621	*TRIM10*	chr6: 30122565	Open Sea	Body	-0.1047	0.0213	2.13E-06	0.465
cg23095517	*SOX13*	chr1: 204093985	Open Sea	Body	-0.0299	0.0062	3.57E-06	0.465
cg16340918	*KIF11*	chr10: 94352700	Island	TSS200	-0.0121	0.0026	5.37E-06	0.465
cg10283879	—	chr3: 195343007	North Shelf	—	-0.0420	0.0090	6.66E-06	0.465
cg02606808	*MAP1B*	chr5 :71403502	Island	1stExon	-0.0112	0.0024	7.09E-06	0.465
cg16688681	*GNA12*	chr7: 2874522	Open Sea	Body	0.1149	0.0248	7.66E-06	0.465
cg26611723	—	chr1: 91173030	South Shore	—	0.0141	0.0030	7.66E-06	0.465
cg07873154	*TRIM10*	chr6 :30122544	Open Sea	Body	-0.1022	0.0222	8.62E-06	0.465
cg11857238	*TRIM10*	chr6 :30122593	Open Sea	Body	-0.1261	0.0281	1.35E-05	0.610
cg05845236	*PCDH9*	chr13: 67804635	Island	TSS200	-0.0161	0.0036	1.46E-05	0.610
cg08094206	*TRIM10*	chr6: 30122523	Open Sea	Body	-0.0977	0.0219	1.51E-05	0.610
cg24597131	*KIAA1026*	chr1: 15354730	Open Sea	Body	-0.0217	0.0049	1.89E-05	0.652
cg09681286	*ABCA3*	chr16: 2334256	Island	Body	0.0282	0.0064	1.94E-05	0.652
cg13474520	*FKTN*	chr9: 108320327	Island	TSS200	-0.0134	0.0031	2.01E-05	0.652
cg23684410	*SIK3*	chr11: 116897558	Open Sea	Body	0.0505	0.0115	2.17E-05	0.657
cg17160660	*MYC*	chr8: 128746896	North Shore	TSS1500	0.0588	0.0136	2.71E-05	0.760
cg01822785	*DDX4*	chr5: 55032703	North Shore	TSS1500	-0.0311	0.0072	2.91E-05	0.760
cg18432864	*SIK3*	chr11: 116969101	Island	TSS200	-0.0301	0.0070	2.97E-05	0.760
cg04233421	*NOTCH4*	chr1: 38679372	Open Sea	—	-0.0472	0.0110	3.18E-05	0.771

The number of subjects included in the two models is n = 176. β value (0–1 scale) represents the difference in methylation for a 1% change in mtDNA content relative to its reference genes. Column headers: Gene = UCSC annotated gene; Position = chromosome and chromosomal position; Localization on CGI= UCSC gene region feature category; Localization on CGI = UCSC relation to CpG islands; β = regression coefficient; SE = standard error for regression coefficient.

Table S4.

Pathways and gene-disease-associations exclusively found for the association with mtDNA content.

Database	Pathway	Term	Association with mtDNA content		
			Bonferroni-p-value (5%)	Count %	Fold Enrichment
Reactome	R-HSA-5663220	RHO GTPases Activate Formins	1.80E-06	4.47	5.34
Reactome	R-HSA-189451	Heme biosynthesis	0.0024	1.34	19.64
Reactome	R-HSA-1222556	ROS. RNS production in phagocytes	0.012	1.79	8.47
Reactome	R-HSA-1236973	Cross-presentation of particulate exogenous antigens (phagosomes)	0.014	4.47	5.34

Table S5.

Pathways and gene-disease-associations exclusively found for the association with $PM_{2.5}$.

Database	Pathway	Term	Association with $PM_{2.5}$ content		
			Bonferroni-p-value (5%)	Count %	Fold Enrichment
Reactome	R-HSA-73728	RNA Polymerase I Promoter Opening	3.78E-07	5.85	14.95
Reactome	R-HSA-2559586	DNA Damage/Telomere Stress Induced Senescence	6.07E-07	5.85	14.27
Reactome	R-HSA-2559582	Senescence-Associated Secretory Phenotype (SASP)	7.38E-07	6.91	10.12
Reactome	R-HSA-171306	Incorporation Of Extended And Processed Telomere End Into Higher Order T-Loop And Associated Protein Structure	2.14E-05	4.79	14.82
Reactome	R-HSA-2559580	Oxidative Stress Induced Senescence	2.72E-05	6.39	8.35
Reactome	R-HSA-3214842	HDMs demethylate histones	0.0039	3.72	11.99
KEGG	hsa04114	Oocyte meiosis	0.01	4.26	7.25
Reactome	R-HSA-5693607	Processing of DNA double-strand break ends	0.023	4.26	6.99
Reactome	R-HSA-5693571	Nonhomologous End-Joining (NHEJ)	0.024	3.72	8.69
Reactome	R-HSA-1538133	G0 and Early G1	0.031	2.66	17.12

Table S6.

Differentially methylated regions significantly associated with $PM_{2.5}$ identified by the DMRcate algorithm.

Gene	Position	Location	no.cpgs	meanbetafc	Stouffer
RPTOR	chr17:78863570-78866579	covers exon(s)	12	0.0068	0.0113
RNF126	chr19:649039-649464	covers exon(s)	3	0.0021	0.0411
UBE2MP1	chr16:34456419-34456429	upstream	2	0.0044	0.0480
VWA1	chr1:1373044-1374310	inside intron	4	0.0034	0.0797
LOC441242	chr7:64894934-64895418	downstream	3	0.0031	0.0802
SNORA71B	chr20:37054880-37056075	promoter	4	0.0021	0.0887
NDRG2	chr14:21526267-21526351	inside intron	2	0.0033	0.0927
AGPAT5	chr8:6470040-6470839	upstream	4	0.0020	0.0959
KDM5B	chr1:202778443-202779497	promoter	5	-00.089	0.0988
CCHCR1	chr6:31116343-31116408	overlaps exon	2	0.0019	0.1019
NFIB	chr9:14314066-14314158	inside intron	2	-0.0029	0.1035
ZNF767P	chr7:149461522-149461763	upstream	2	0.0022	0.1106
TRIP10	chr19:6670865-6671045	upstream	2	0.0023	0.1115
ATP5EP2	chr13:28519319-28519388	overlaps 5'	2	0.0031	0.1149
USP28	chr11:114127924-114127960	upstream	2	0.0024	0.1162
HDAC4	chr2:240196769-240197131	inside intron	4	0.0024	0.1162
PEX11G	chr19:7676731-7676786	upstream	2	0.0026	0.1225
TELO2	chr16:1444138-1444162	upstream	2	0.0023	0.1231
CFAP46	chr10:134699444-134699744	downstream	3	0.0017	0.1238
ADGRB1	chr8:143545940-143545949	inside exon	2	0.0042	0.1308
FIBCD1	chr9:133797391-133797653	inside intron	2	0.0029	0.1332
PIK3R3	chr1:46639632-46639752	upstream	2	0.0025	0.1338
CRAMP1	chr16:1682412-1682446	inside intron	2	0.0014	0.1358
TET1	chr10:70184347-70184359	upstream	2	0.0020	0.1367
CKMT2	chr5:80550841-80550860	inside exon	2	0.0024	0.1379
MRGPRG	chr11:3239402-3240399	overlaps 5'	8	0.0024	0.1388
TPST2	chr22:26961237-26961465	overlaps 5'	3	0.0023	0.1397
WIZ	chr19:15559494-15559628	upstream	2	0.0022	0.1447
ZMYND12	chr1:43204163-43204185	upstream	2	0.00022	0.1456
ERRFI1	chr1:8383169-8383171	upstream	2	0.0031	0.1457
FLJ13224	chr12:31272112-31272119	upstream	3	0.0167	0.1482
PMFBP1	chr16:72830034-72830284	upstream	3	0.0008	0.1489
WDR45B	chr17:80553970-80554007	downstream	2	-0.0054	0.1518
MOSPD3	chr7:100203350-100203651	upstream	3	0.0001	0.1526
TNFRSF4	chr1:1146777-1146903	inside exon	2	0.0027	0.1536
PSG11	chr19:43530536-43530623	inside exon	2	0.0025	0.1552
GLIS2	chr16:3812613-3812688	upstream	3	0.0016	0.1557

ANKRD33	chr12:52281482-52282079	overlaps 5'	8	-0.0041	0.1565
NRXN2	chr11:64270224-64270338	downstream	3	0.0038	0.1577
FOXA3	chr19:46288852-46288989	upstream	2	0.0018	0.1583
SCART1	chr10:135270925-135271052	upstream	3	0.0023	0.1598
PTPRF	chr1:44056769-44057099	inside exon	4	0.0026	0.1607
AATK	chr17:79077169-79077330	downstream	2	0.0019	0.1613
CHST12	chr7:2472814-2473529	inside exon	4	0.0016	0.1625
ACSBG2	chr19:6271960-6271968	downstream	2	0.0025	0.1627
RAD54L	chr1:46751642-46751975	downstream	2	0.0017	0.1660
OR52H1	chr11:5567996-5568008	promoter	2	0.0047	0.1677
MAD1L1	chr7:2260901-2260999	inside intron	2	0.0018	0.1679
IRX1	chr5:4941328-4941407	downstream	2	0.0021	0.1686
GRTP1	chr13:114071457-114071944	upstream	3	0.0033	0.1686
GBX2	chr2:236867685-236867836	downstream	2	0.0019	0.1713
NDUFS6	chr5:2175315-2175329	downstream	2	0.0020	0.1718
MTG1	chr10:135202522-135203200	upstream	7	0.0069	0.1730
RAB17	chr2:239037049-239037533	upstream	4	0.0071	0.1731
ARHGEF26	chr3:153839115-153839287	overlaps 5'	4	-0.0013	0.1746
FGFRL1	chr4:987561-987906	upstream	3	-0.0002	0.1757
AKAP13	chr15:86233214-86233236	inside intron	3	-0.0012	0.1766
G0S2	chr1:209847618-209847729	promoter	2	0.0031	0.1767
PXDN	chr2:1681098-1681494	inside intron	3	0.0021	0.1786
MBP	chr18:77560089-77560291	upstream	2	0.0033	0.1832
MAP4K3	chr2:39892450-39892612	upstream	3	-0.0027	0.1835
SGSH	chr17:78313648-78314035	upstream	4	0.0024	0.1847
MYADM	chr19:54312855-54313004	upstream	3	0.0034	0.1859
MRPL28	chr16:426086-426799	upstream	4	0.0011	0.1862
KXD1	chr19:18549689-18550061	upstream	3	0.0057	0.1864
PI3	chr20:43743301-43743511	upstream	2	0.0023	0.1892
ARFGAP1	chr20:61915437-61916072	overlaps 5'	4	0.0030	0.1930
IMPAD1	chr8:58173162-58173601	upstream	2	0.0019	0.1965
OR6C75	chr12:55758598-55758751	promoter	2	0.0047	0.1995
NDUFA3	chr19:54603567-54604187	promoter	4	0.0023	0.2000

Column headers: Gene = UCSC annotated gene; Position = chromosome and chromosomal position; Location= UCSC gene region feature category; no.cpgs = number of CpGs constituting the differentially methylated region; meanbetafc = mean ß fold change; Stouffer = Stouffer's Z-score.

Table S7.

CpG sites constituting the differentially methylated region annotated to *RPTOR* as identified by DMRcate for the association with $PM_{2.5}$. The two CpG sites which reached a significant FDR value < 0.2 are in bold.

ID	Position	Localization on CGI	Localization on gene	betafc	FDR
cg16018154	78863570	Island	Body	0.0100	0.29
cg25902229	78864088	South Shore	Body	0.0100	0.24
cg05774614	78864539	South Shore	Body	**0.0088**	**0.12**
cg10035831	78865087	South Shore	Body	0.0023	0.27
cg22636722	78865263	South Shore	Body	0.0053	0.25
cg00704970	78865368	South Shore	Body	0.0093	0.24
cg03502601	78865373	South Shore	Body	0.0076	0.26
cg09803959	78865514	South Shore	Body	0.0056	0.31
cg24207068	78865662	South Shore	Body	0.0091	0.21
cg04658243	78865755	South Shore	Body	**0.0072**	**0.18**
cg11476241	78866235	North Shelf	Body	0.0051	0.29
cg24327522	78866579	North Shelf	Body	0.0012	0.50

Position = position on CHR 17; Localization on CGI= UCSC gene region feature category; Localization on CGI = UCSC relation to CpG islands; betafc = ß fold change of the individual CpG site.

Table S8.

CpG sites constituting the differentially methylated region annotated to *VWA1* as identified by DMRcate for the association with $PM_{2.5}$. The two CpG sites which reached a significant FDR value < 0.2 are in bold.

ID	Position	Localization on CGI	Localization on gene	betafc	FDR
cg26798702	1373044	North Shore	Body, 3'UTR	0.00098	0.647
cg09843049	1373316	North Shore	Body, 3'UTR	**0.00586**	**0.018**
cg20255272	1373678	North Shore	Body, 3'UTR	**0.00550**	**0.152**
cg13897675	1374310	North Shore	Body, 3'UTR	0.00139	0.472

Position = position on CHR 1; Localization on CGI= UCSC gene region feature category; Localization on CGI = UCSC relation to CpG islands; betafc = ß fold change of the individual CpG site.

Table S9.

CpG sites constituting the differentially methylated region annotated to *PSG11* as identified by DMRcate for the association with $PM_{2.5}$. The CpG site which reached a significant FDR value < 0.2 is in bold.

ID	Position	Localization on CGI	Localization on gene	betafc	FDR
cg26140182	**1373044**	**Open Sea**	**5'UTR, 1stExon**	**0.0044**	**0.014**
cg07920195	1374310	Open Sea	5'UTR, 1stExon	0.0005	0.77

Position = position on CHR 1; Localization on CGI= UCSC gene region feature category; Localization on CGI = UCSC relation to CpG islands; betafc = ß fold change of the individual CpG site

Table S10.

CpG sites constituting the candidate differentially methylated region annotated to *STIL* identified by bumphunter for the association with $PM_{2.5}$ and mtDNA content.

ID	Position	Localization on CGI	Localization on gene	$PM_{2.5}$		mtDNA content	
				betafc	*p*-value	betafc	*p*-value
cg10715383	47800167	Island	Body	-0.0018	0.87	0.0891	0.41
cg22532312	47800409	Island	Body	0.0001	0.80	-0.0041	0.47

Position = position on CHR 1; Localization on CGI= UCSC gene region feature category; Localization on CGI = UCSC relation to CpG islands; betafc = ß fold change of the individual CpG site; *p*-value=marginal p-value for each genomic location

Table S11.

CpG sites constituting the candidate differentially methylated region annotated to *SPRN* identified by bumphunter for the association with mtDNA content.

ID	Position	Localization on CGI	Localization on gene	betacf	*p*-value
cg13315147	135341528	Island	Body	0.171	0.015
cg19469447	135341870	Island	Body	0.149	0.020
cg00321709	135341933	Island	Body	0.161	0.024
cg10862468	135342218	Island	Body	0.156	0.025
cg25330361	135342413	Island	Body	0.044	0.037
cg23400446	135342560	Island	Body	0.142	0.020
cg24530264	135342620	South Shore	Body	0.096	0.028
cg18984983	135342936	South Shore	Body	0.098	0.039
cg05194426	135343193	South Shore	Body	0.217	0.002
cg11445109	135343248	South Shore	Body	0.174	0.008
cg27214960	135343280	South Shore	Body	0.142	0.007

Position = position on CHR 1; Localization on CGI= UCSC gene region feature category; Localization on CGI = UCSC relation to CpG islands; betafc = ß fold change of the individual CpG site; p-value=marginal p-value for each genomic location

Table S12.

Top 20 differentially methylated regions (DMRs) highest ranked according to FWER for the association with $PM_{2.5}$ identified by the bumphunter algorithm. DMRs with an overlap within the top 20 found for the association with mtDNA content are shown in bold.

Gene	Coordinates	Location	no.cpgs	meanbetafc	FWER
MOG	chr6:29648161-29649084	downstream	22	-0.007	0.26
STIL	**chr1:47800167-47800409**	**upstream**	**2**	**0.022**	**0.35**
LY6G5C	chr6:31650735-31651362	covers exon(s)	21	-0.007	0.41
HLA-DRB6	**chr6:32551749-32552670**	**overlaps 5'**	**14**	**0.006**	**0.57**
SCAMP1	chr5:77146796-77147141	upstream	3	-0.014	0.66
LINC02139	chr16:85242230-85242550	downstream	4	-0.012	0.68
HLA-DPB1	chr6:33047944-33048879	overlaps 5'	21	0.004	0.72
CERS3	chr15:101084507-101085341	overlaps 5'	8	0.009	0.79
KRTCAP3	chr2:27665017-27665711	overlaps 5'	11	-0.007	0.83
CCKBR	**chr11:6291339-6292896**	**covers**	**12**	**0.005**	**0.84**
BCL11B	chr14:99736387-99736387	inside intron	1	0.016	0.91
CLIC1	chr6:31712014-31712195	upstream	2	-0.013	0.93
CRISP2	chr6:49681178-49681774	overlaps 5'	9	-0.007	0.93
SCMH1	chr1:41707370-41707379	inside intron	2	-0.013	0.94
CCDC185	chr1:223566127-223567173	overlaps 5'	11	0.006	0.94
CPT1B	**chr22:51016501-51017162**	**overlaps 5'**	**13**	**-0.004**	**0.95**
HOXA5	chr7:27183133-27183950	overlaps 5'	19	-0.003	0.95
PRR34	chr22:46449430-46450114	overlaps 5'	11	-0.006	0.95
MOB2	chr11:1859381-1859510	upstream	2	-0.012	0.96
BICDL1	chr12:120554733-120555025	downstream	2	0.012	0.97

Column headers: Gene = UCSC annotated gene; Position = chromosome and chromosomal position; Location= UCSC gene region feature category; no.cpgs = number of CpGs constituting the differentially methylated region; meanbetafc = mean ß fold change; FWER = Family wise error rate.

Table S13.

Top 20 differentially methylated regions (DMRs) highest ranked according to FWER for the association with mtDNA content identified by the bumphunter algorithm. DMRs with an overlap within the top 20 found for the association with $PM_{2.5}$ are shown in bold.

Gene	Coordinates	Location	no.cpgs	meanbetafc	FWER
STIL	**chr1:47800167-47800409**	**upstream**	**2**	**0.504**	**0.08**
SPRN	chr10:135341528-135343280	upstream	11	0.145	0.08
LINC00662	chr19:29217858-29218774	upstream	7	0.148	0.21
PSCA	chr8:143751447-143751801	overlaps 5'	5	0.136	0.39
PKN1	chr19:14544684-14544684	upstream	1	-0.258	0.40
RUFY1	chr5:178986131-178986906	overlaps 5'	9	-0.083	0.83
HLA-DQB2	chr6:32729118-32729823	covers exon(s)	21	-0.036	0.84
TBCD	chr17:80889693-80890438	inside intron	4	-0.109	0.84
GNA11	chr19:3062217-3062219	upstream	2	0.140	0.86
HLA-DQB1	chr6:32632568-32633163	overlaps 5'	11	-0.063	0.89
MUC4	chr3:195488725-195490309	covers exon(s)	11	-0.058	0.93
HCG27	chr6:31148332-31148748	upstream	14	0.041	0.95
ALKBH7	chr19:6373461-6373627	inside intron	2	0.111	0.99
TCP11L2	chr12:106696891-106697297	overlaps 5'	3	0.090	1.00
HLA-DRB6	**chr6:32551749-32552246**	**overlaps 5'**	**10**	**-0.050**	**1.00**
PRDM8	chr4:81117647-81119473	overlaps 5'	11	-0.043	1.00
RHCG	chr15:89959984-89960743	downstream	7	-0.065	1.00
CPT1B	**chr22:51016501-51017151**	**overlaps 5'**	**12**	**0.035**	**1.00**
AURKC	chr19:57741988-57742444	overlaps 5'	10	0.041	1.00
CCKBR	**chr11:6291447-6292896**	**covers**	**11**	**0.037**	**1.00**

Column headers: Gene = UCSC annotated gene; Position = chromosome and chromosomal position; Location= UCSC gene region feature category; no.cpgs = number of CpGs constituting the differentially methylated region; meanbetafc = mean ß fold change; FWER = Family wise error rate.

Table S14.

CpG sites constituting the candidate differentially methylated region annotated to *HLA-DRB6* identified by bumphunter for the association with $PM_{2.5}$ and mtDNA content.

ID	Position	Localization on CGI	Localization on gene	$PM_{2.5}$		mtDNA content	
				betafc	*p*-value	betafc	*p*-value
cg11404906	32551749	North Shore	Body	0.0080	0.44	-0.060	0.61
cg15568074	32551949	Island	Body	0.0051	0.63	-0.111	0.37
cg24242384	32551954	Island	Body	0.0028	0.59	-0.019	0.75
cg17316649	32552016	Island	Body	0.0082	0.07	-0.032	0.52
cg08578320	32552039	Island	Body	0.0087	0.30	-0.016	0.86
cg09139047	32552042	Island	Body	0.0055	0.60	-0.038	0.73
cg15982117	32552106	Island	Body	0.0122	0.15	-0.055	0.56
cg16514085	32552152	Island	Body	0.0063	0.59	-0.153	0.21
cg14645244	32552205	Island	Body	0.0059	0.55	-0.049	0.65
cg00211215	32552246	Island	Body	0.0031	0.84	-0.293	0.07
cg09949906	32552350	South Shore	Body	0.0082	0.42	—	—
cg26036029	32552443	South Shore	Body	0.0048	0.04	—	—
cg10632894	32552453	South Shore	Body	0.0081	0.02	—	—
cg17480035	32552670	South Shore	Body	-0.0069	0.56	—	—

Position = position on CHR 6; Localization on CGI= UCSC gene region feature category; Localization on CGI = UCSC relation to CpG islands; betafc = ß fold change of the individual CpG site; *p*-value=marginal p-value for each genomic location.

Table S15.

CpG sites constituting the candidate differentially methylated region annotated to *CPT1B* identified by bumphunter for the association with $PM_{2.5}$ and mtDNA content.

ID	Position	Localization on CGI	Localization on gene	$PM_{2.5}$		mtDNA content	
				betafc	*p*-value	betafc	*p*-value
cg08260245	51016501	Island	5'UTR, 1st exon	-0.0061	0.05	0.037	0.22
cg10490842	51016604	Island	5'UTR, TSS200	-0.0037	0.08	0.026	0.21
cg19112186	51016638	Island	5'UTR, TSS200	-0.0061	0.10	0.049	0.16
cg10770023	51016644	Island	5'UTR, TSS200	-0.0033	0.29	0.016	0.63
cg05156901	51016646	Island	5'UTR, TSS200	-0.0059	0.13	0.058	0.12
cg24363820	51016703	Island	5'UTR, TSS200	-0.0071	0.04	0.049	0.14
cg00047287	51016899	Island	TSS200, TSS1500	-0.0021	0.27	0.017	0.37
cg06530441	51016950	Island	TSS200, TSS1500	-0.0079	0.02	0.041	0.20
cg27502912	51017001	Island	1stExon, TSS1500	-0.0035	0.21	0.018	0.48
cg00270625	51017019	Island	1stExon, TSS1500	-0.0084	0.04	0.078	0.06
cg00983520	51017067	South Shore	1stExon, TSS1500	-0.0042	0.08	0.033	0.16
cg01081346	51017151	South Shore	TSS200, TSS1500	-0.0033	0.21	0.037	0.15
cg16386697	51017162	South Shore	TSS200, TSS1500	-0.0038	0.08	—	—

Position = position on CHR 22; Localization on CGI= UCSC gene region feature category; Localization on CGI = UCSC relation to CpG islands; betafc = ß fold change of the individual CpG site; *p*-value=marginal p-value for each genomic location.

Table S16.

CpG sites constituting the candidate differentially methylated region annotated to *CCKBR* identified by bumphunter for the association with $PM_{2.5}$ and mtDNA content.

ID	Position	Localization on CGI	Localization on gene	$PM_{2.5}$		mtDNA content	
				betafc	*p*-value	betafc	*p*-value
cg11812625	6291339	Island	Body	0.0051	0.06	—	—
cg27443416	6291447	Island	Body	0.0038	0.17	0.028	0.33
cg21112490	6291549	Island	Body	0.0023	0.14	0.016	0.32
cg13580265	6291625	North Shore	Body	0.0060	0.06	0.040	0.26
cg19364351	6291879	North Shore	Body	0.0066	0.05	0.047	0.18
cg04467334	6291927	North Shore	Body	0.0064	0.04	0.028	0.41
cg26313599	6291951	North Shore	Body	0.0072	0.03	0.020	0.57
cg04585209	6292311	Island	Body	0.0084	0.07	0.065	0.16
cg18706028	6292490	Island	Body	0.0102	0.03	0.051	0.29
cg26700447	6292511	Island	Body	0.0078	0.06	0.064	0.13
cg09656511	6292615	Island	Body	0.0023	0.06	0.023	0.09
cg25740457	6292896	South Shore	3'UTR	0.0029	0.20	0.039	0.10

Position = position on CHR 11; Localization on CGI= UCSC gene region feature category; Localization on CGI = UCSC relation to CpG islands; betafc = ß fold change of the individual CpG site; *p*-value=marginal p-value for each genomic location.

Supplementary figures

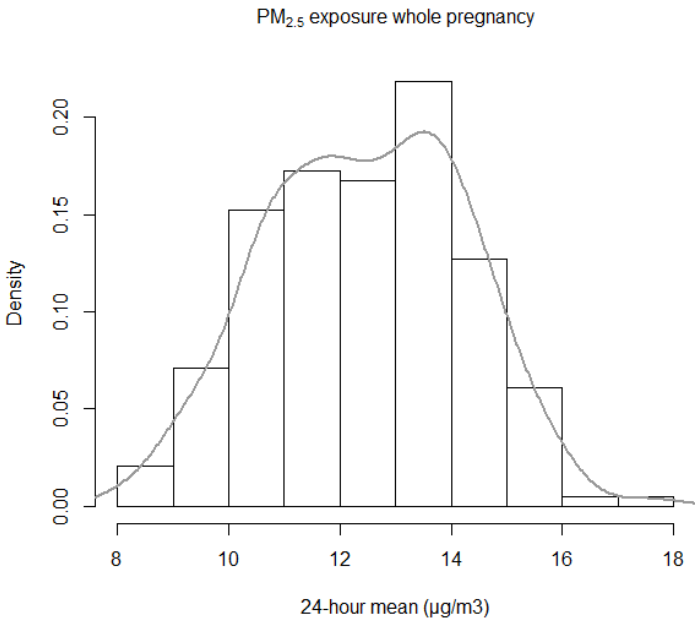


Figure S1. Distribution of the 24-hour mean of $PM_{2.5}$ exposure during the course of the entire pregnancy in n= 187 mothers from the ENVIR*ON*AGE cohort.

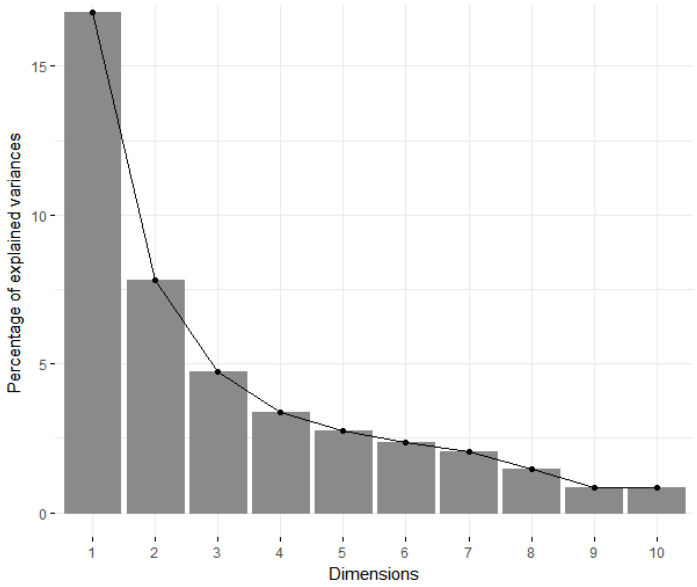
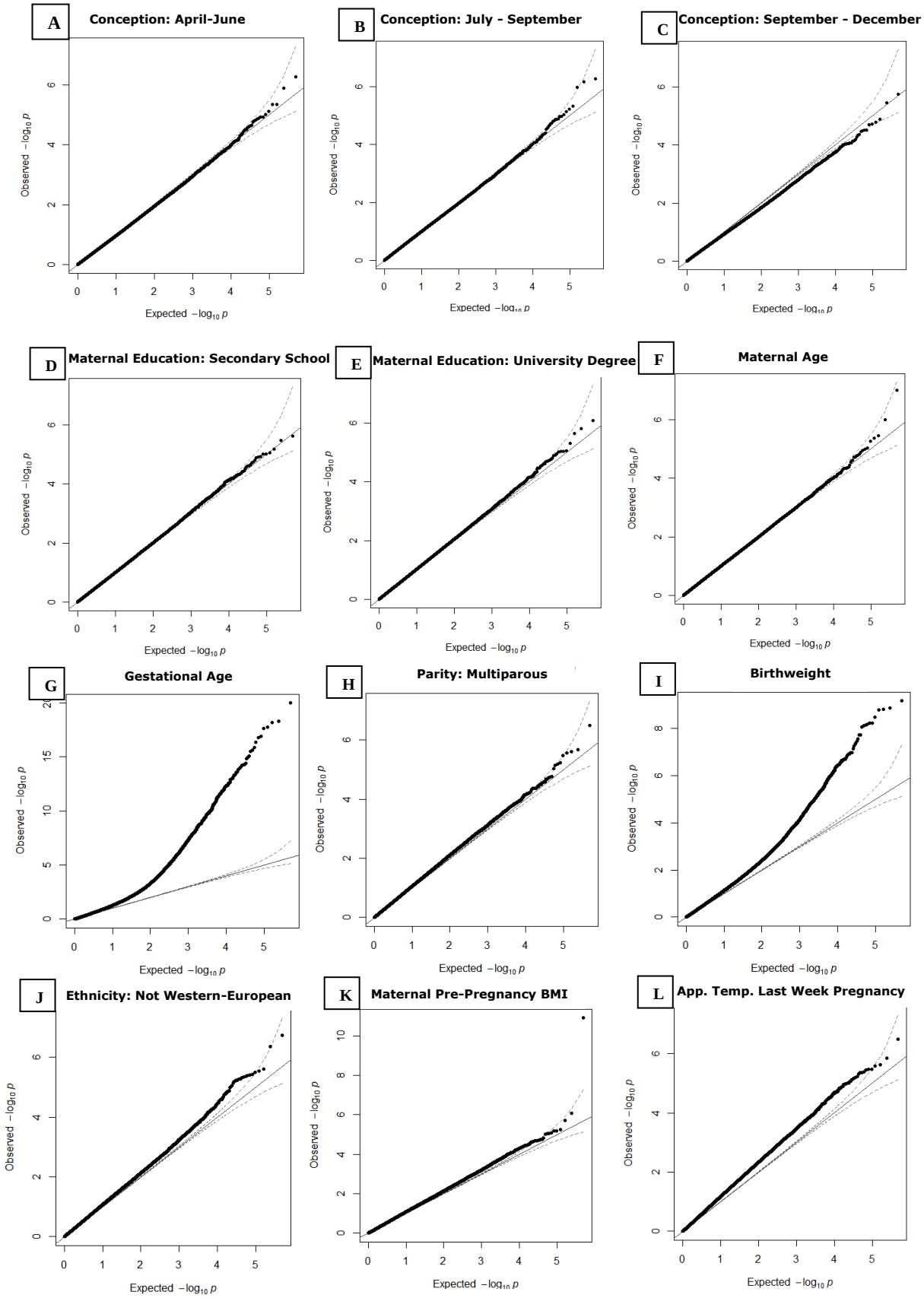
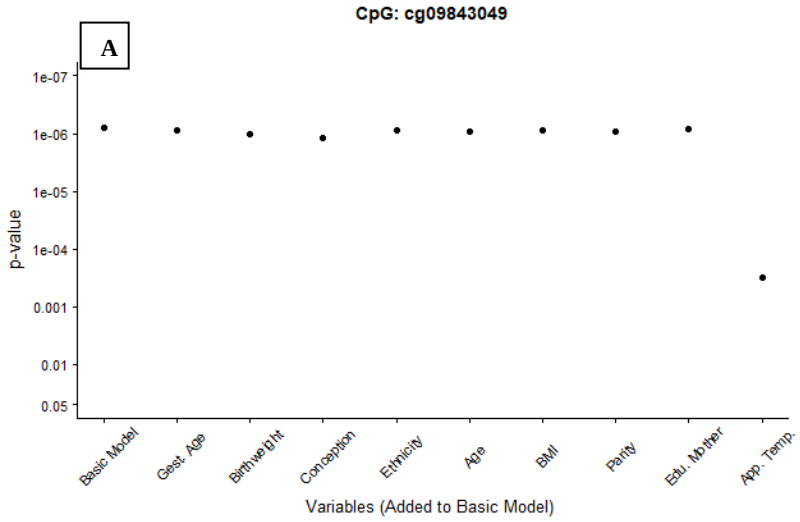
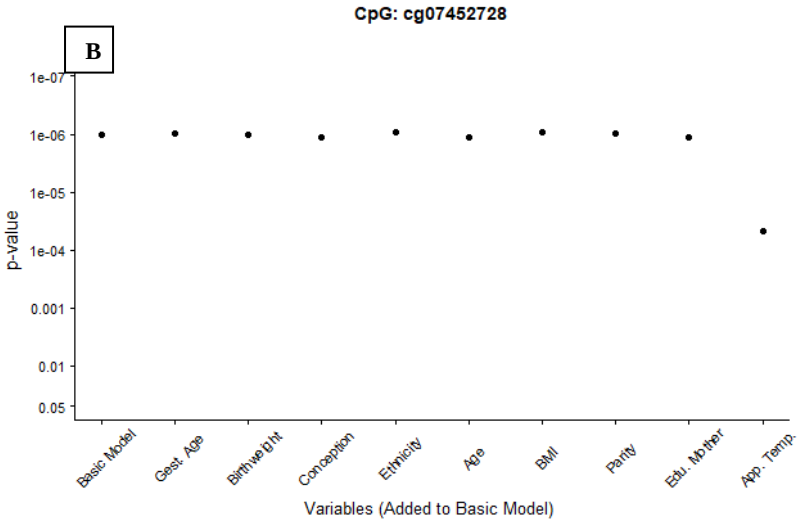


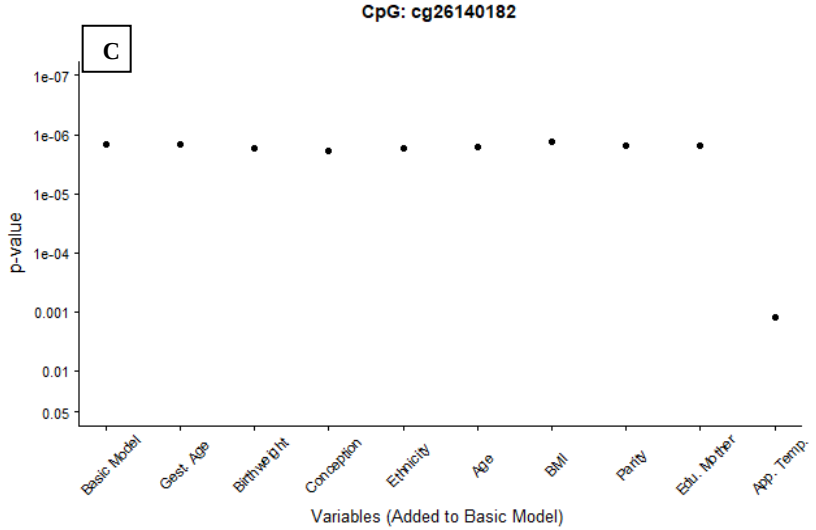
Figure S2. Eigenvalues of the top 10 principal components which explain cumulatively ca. 43% of the variance in the Beta-values.

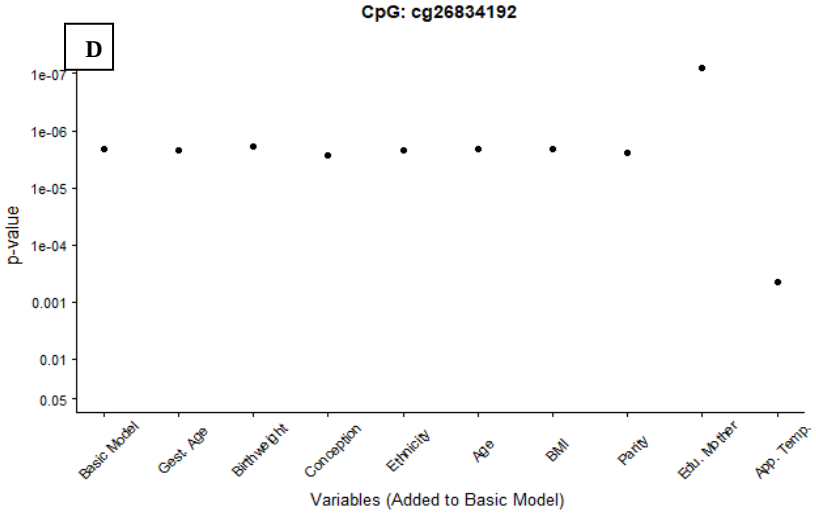


Figures S3 A-L. QQ plots of the observed vs the expected *p*-values for potentially confounding variables in a multiple linear mixed regression model.


A
CpG: cg09843049
p-value
1e-07
1e-06
1e-05
1e-04
0.001
0.01
0.05
Basic Model
Gest. Age
Birthweight
Conception
Ethnicity
Age
BMI
Parity
Edu. Mother
App. Temp.
Variables (Added to Basic Model)


B
CpG: cg07452728
p-value
1e-07
1e-06
1e-05
1e-04
0.001
0.01
0.05
Basic Model
Gest. Age
Birthweight
Conception
Ethnicity
Age
BMI
Parity
Edu. Mother
App. Temp.
Variables (Added to Basic Model)


C
CpG: cg26140182
p-value
1e-07
1e-06
1e-05
1e-04
0.001
0.01
0.05
Basic Model
Gest. Age
Birthweight
Conception
Ethnicity
Age
BMI
Parity
Edu. Mother
App. Temp.
Variables (Added to Basic Model)

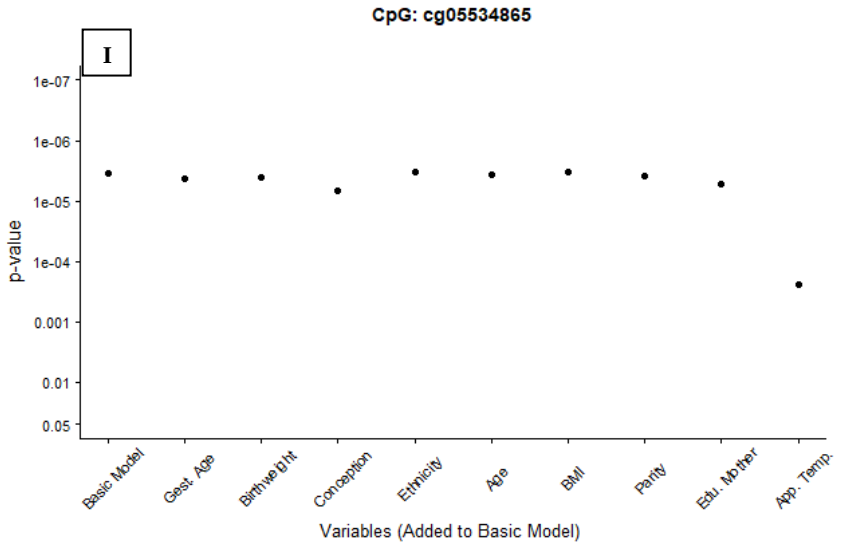

D
CpG: cg26834192
p-value
1e-07
1e-06
1e-05
1e-04
0.001
0.01
0.05
Basic Model
Gest. Age
Birthweight
Conception
Ethnicity
Age
BMI
Parity
Edu. Mother
App. Temp.
Variables (Added to Basic Model)

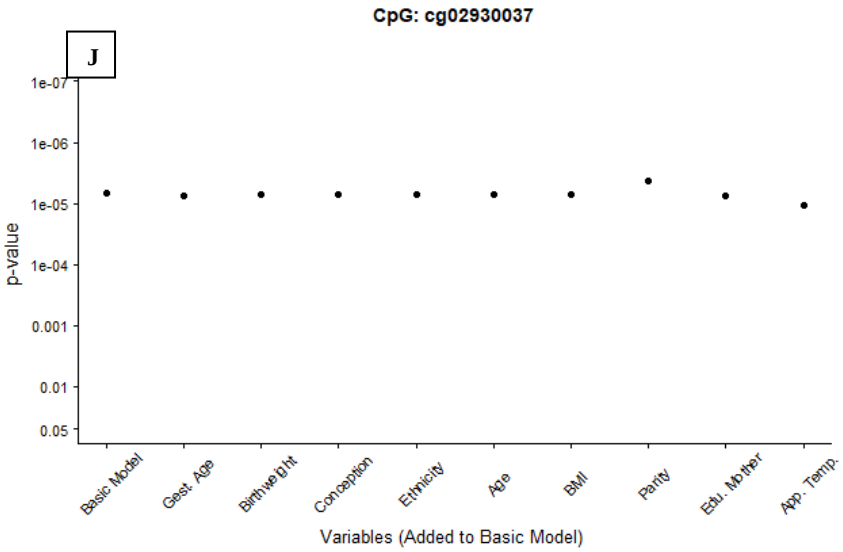
E
CpG: cg03804903
p-value
1e-07
1e-06
1e-05
1e-04
0.001
0.01
0.05
Basic Model
Gest. Age
Birthweight
Conception
Ethnicity
Age
BMI
Parity
Edu. Mother
App. Temp.
Variables (Added to Basic Model)

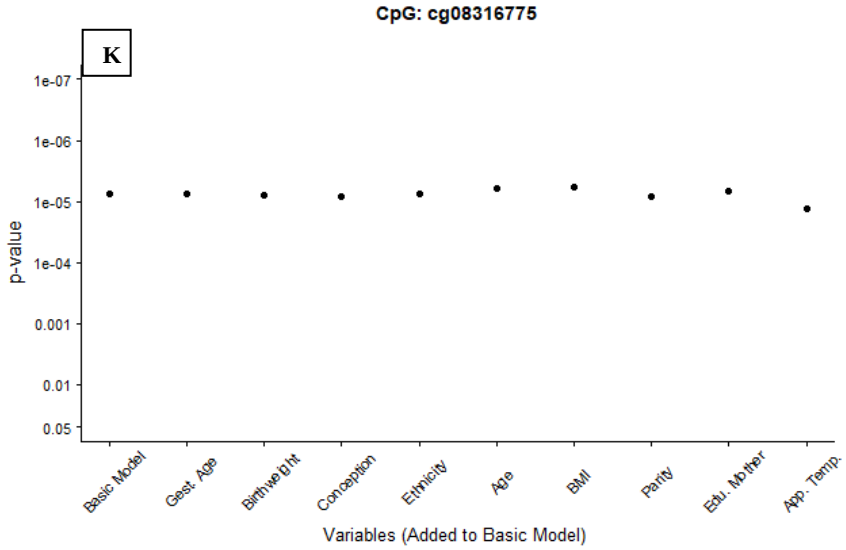
F
CpG: cg05600804
p-value
1e-07
1e-06
1e-05
1e-04
0.001
0.01
0.05
Basic Model
Gest. Age
Birthweight
Conception
Ethnicity
Age
BMI
Parity
Edu. Mother
App. Temp.
Variables (Added to Basic Model)

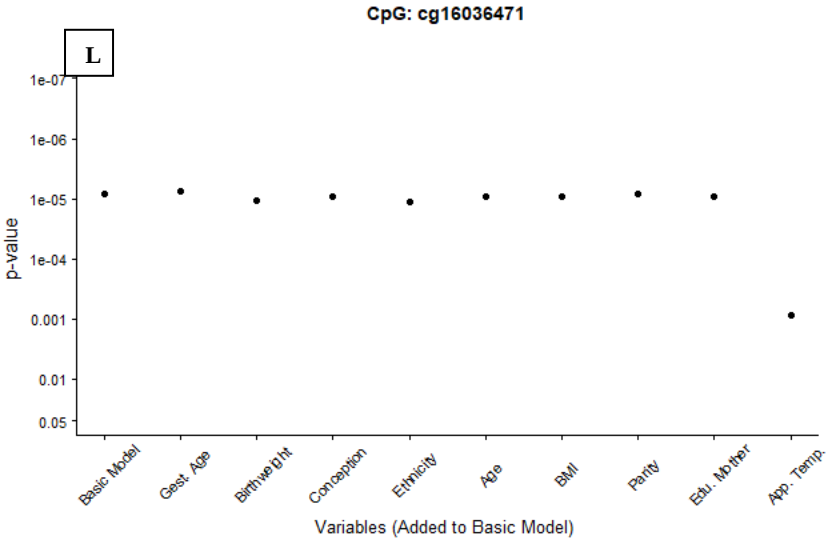
G
CpG: cg14173815
p-value
1e-07
1e-06
1e-05
1e-04
0.001
0.01
0.05
Basic Model
Gest. Age
Birthweight
Conception
Ethnicity
Age
BMI
Parity
Edu. Mother
App. Temp.
Variables (Added to Basic Model)

H
CpG: cg00308185
p-value
1e-07
1e-06
1e-05
1e-04
0.001
0.01
0.05
Basic Model
Gest. Age
Birthweight
Conception
Ethnicity
Age
BMI
Parity
Edu. Mother
App. Temp.
Variables (Added to Basic Model)


I
CpG: cg05534865
p-value
1e-07
1e-06
1e-05
1e-04
0.001
0.01
0.05
Basic Model
Gest. Age
Birthweight
Conception
Ethnicity
Age
BMI
Parity
Edu. Mother
App. Temp.
Variables (Added to Basic Model)


J
CpG: cg02930037
p-value
1e-07
1e-06
1e-05
1e-04
0.001
0.01
0.05
Basic Model
Gest. Age
Birthweight
Conception
Ethnicity
Age
BMI
Parity
Edu. Mother
App. Temp.
Variables (Added to Basic Model)


K
CpG: cg08316775
p-value
1e-07
1e-06
1e-05
1e-04
0.001
0.01
0.05
Basic Model
Gest. Age
Birthweight
Conception
Ethnicity
Age
BMI
Parity
Edu. Mother
App. Temp.
Variables (Added to Basic Model)


L
CpG: cg16036471
p-value
1e-07
1e-06
1e-05
1e-04
0.001
0.01
0.05
Basic Model
Gest. Age
Birthweight
Conception
Ethnicity
Age
BMI
Parity
Edu. Mother
App. Temp.
Variables (Added to Basic Model)

M
CpG: cg17448109
p-value
1e-07
1e-06
1e-05
1e-04
0.001
0.01
0.05
Basic Model
Gest. Age
Birthweight
Conception
Ethnicity
Age
BMI
Parity
Edu. Mother
App. Temp.
Variables (Added to Basic Model)

N
CpG: cg27327785
p-value
1e-07
1e-06
1e-05
1e-04
0.001
0.01
0.05
Basic Model
Gest. Age
Birthweight
Conception
Ethnicity
Age
BMI
Parity
Edu. Mother
App. Temp.
Variables (Added to Basic Model)

O
CpG: cg14554618
p-value
1e-07
1e-06
1e-05
1e-04
0.001
0.01
0.05
Basic Model
Gest. Age
Birthweight
Conception
Ethnicity
Age
BMI
Parity
Edu. Mother
App. Temp.
Variables (Added to Basic Model)

P
CpG: cg04283429
p-value
1e-07
1e-06
1e-05
1e-04
0.001
0.01
0.05
Basic Model
Gest. Age
Birthweight
Conception
Ethnicity
Age
BMI
Parity
Edu. Mother
App. Temp.
Variables (Added to Basic Model)

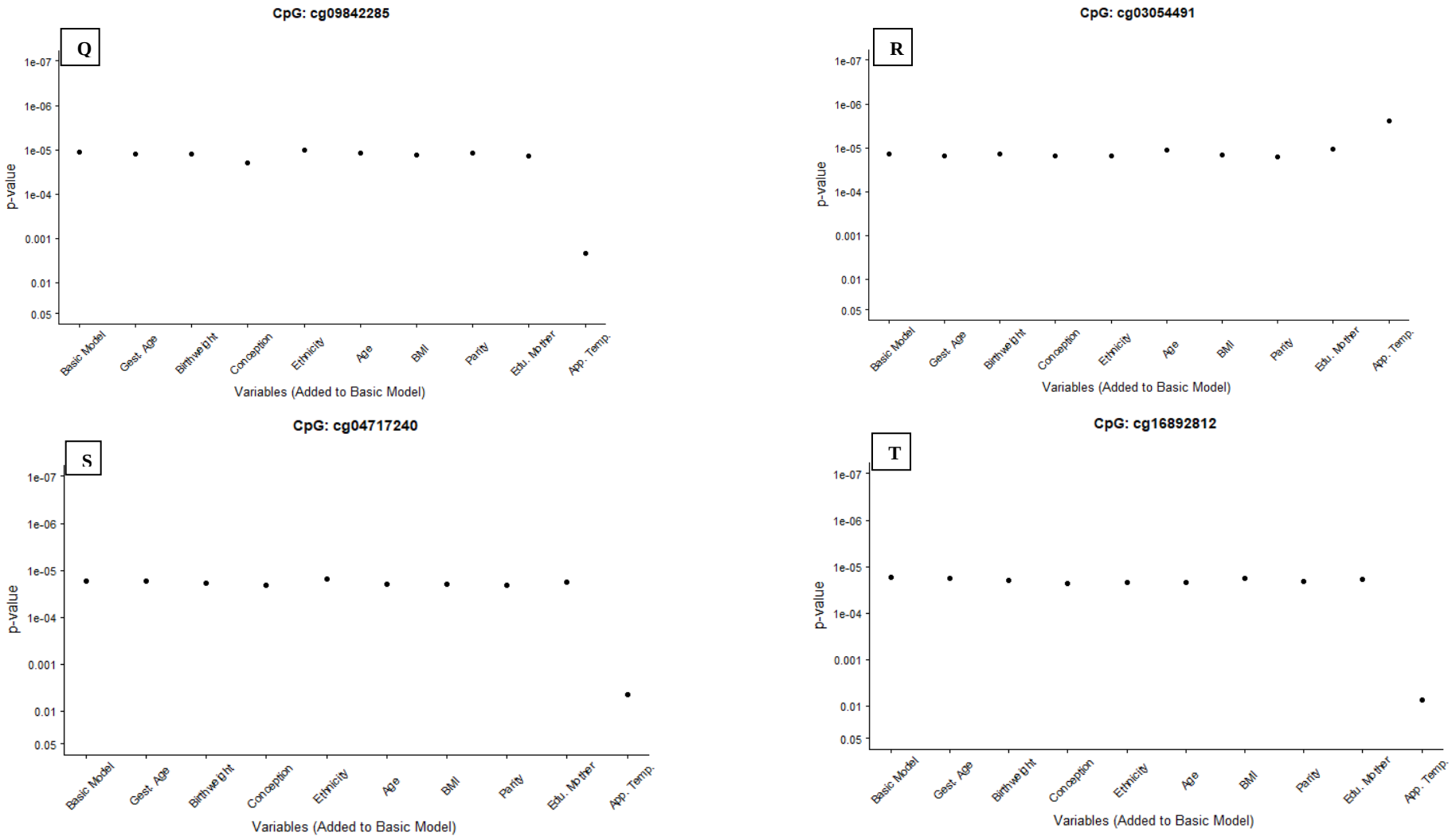


Figure S4 A-T. Differences in *p*-values for the top 20 CpGs according to *p*-values for the association with $PM_{2.5}$ in a multiple linear mixed regression model when the potentially confounding variables are added one at a time to the basic model.

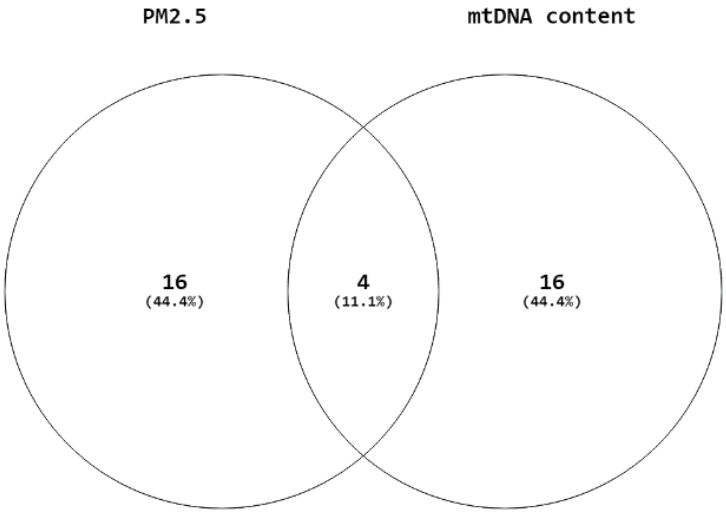


Figure S5. Overlap in genes annotated to the top 20 DMRs according to FWER for the association with $PM_{2.5}$ and mtDNA content found by the bumphunter algorithm.

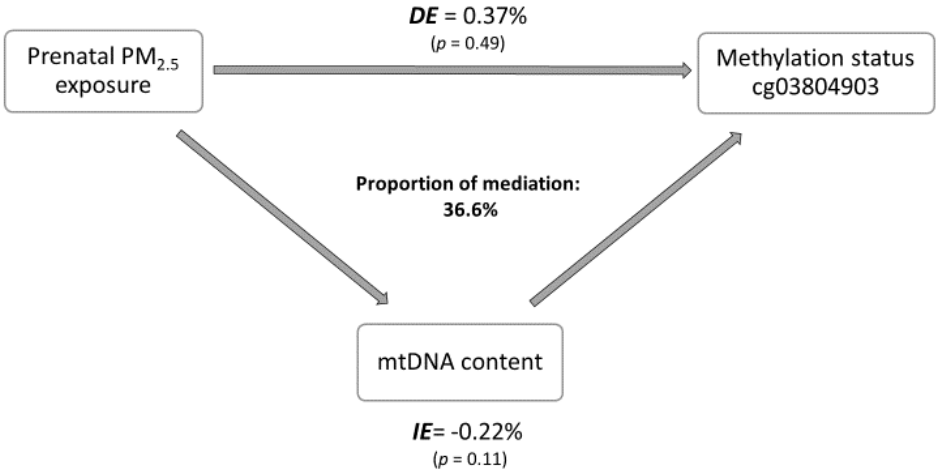


Figure S6. Estimated proportion of the association between a 10 μg/m^3 increment in $PM_{2.5}$ exposure during the entire pregnancy and methylation status of CpG site cg03804903 mediated through mtDNA content. Additionally, the estimates of the indirect effects (IE), the estimates of the direct effect (DE), and the proportion of mediation (IE/DE+IE) are shown. The model was adjusted for chip, position on the chip, white blood cell composition, platelet count, maternal smoking status, age, education, pre-pregnancy BMI, newborns' sex, ethnicity, gestational age, season of conception and apparent temperature in the last week of pregnancy.

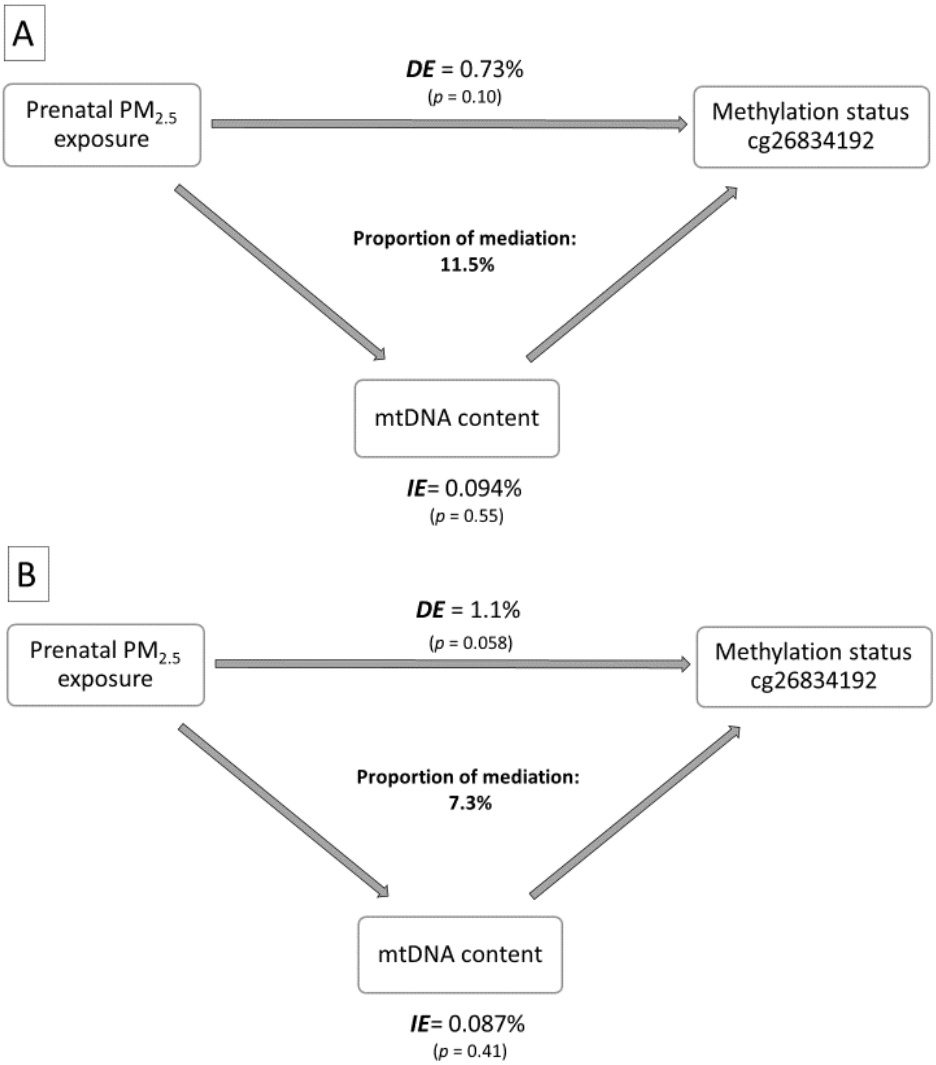


Figure S7. Estimated proportion of the association between a 10 μg/m^3 increment in $PM_{2.5}$ exposure during the entire pregnancy and methylation status of CpG site cg26834192 through mtDNA content. in a model adjusted for (A) chip, position on the chip, white blood cell composition, platelet count, maternal smoking status and newborns' sex and gestational age and (B) additionally for, maternal age, education, pre-pregnancy BMI, newborns' ethnicity, season of conception and apparent temperature in the last week of pregnancy. The estimates of the indirect effects (IE), the estimates of the direct effect (DE), and the proportion of mediation (IE/DE+IE) are also shown. The model was adjusted for chip, position on the chip, white blood cell composition, platelet count, maternal smoking status and newborns' sex and gestational age.

Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:
**Epigenome-wide DNA methylation and mitochondrial DNA content in relation to in utero PM_{2.5} exposure

**

Richting: **Master of Biomedical Sciences-Environmental Health Sciences**
Jaar: **2018**

in alle mogelijke mediaformaten, - bestaande en in de toekomst te ontwikkelen - , aan de Universiteit Hasselt.

Niet tegenstaand deze toekenning van het auteursrecht aan de Universiteit Hasselt behoud ik als auteur het recht om de eindverhandeling, - in zijn geheel of gedeeltelijk -, vrij te reproduceren, (her)publiceren of distribueren zonder de toelating te moeten verkrijgen van de Universiteit Hasselt.

Ik bevestig dat de eindverhandeling mijn origineel werk is, en dat ik het recht heb om de rechten te verlenen die in deze overeenkomst worden beschreven. Ik verklaar tevens dat de eindverhandeling, naar mijn weten, het auteursrecht van anderen niet overtreedt.

Ik verklaar tevens dat ik voor het materiaal in de eindverhandeling dat beschermd wordt door het auteursrecht, de nodige toelatingen heb verkregen zodat ik deze ook aan de Universiteit Hasselt kan overdragen en dat dit duidelijk in de tekst en inhoud van de eindverhandeling werd genotificeerd.

Universiteit Hasselt zal mij als auteur(s) van de eindverhandeling identificeren en zal geen wijzigingen aanbrengen aan de eindverhandeling, uitgezonderd deze toegelaten door deze overeenkomst.

Voor akkoord,

Reimann, Brigitte

Datum: **7/06/2018**