Prenatal and postnatal exposure to tobacco smoking and molecular signatures in children

Marta Vives Usano

TESI DOCTORAL UPF / 2018

DIRECTORES DE LA TESI

Dra. Mariona Bustamente Pineda (ISGlobal – Childhood and Enviornment Program)

Dra. Eulàlia Martí Puig (CRG – Bioinformatics and Genomics/UB – Dpt. Biomedical Sciences)

DEPARTAMENT DE CIÈNCIES EXPERIMENTALS I DE LA SALUT

Universitat Pompeu Fabra Barcelona

Als padrins, Montserrat i Abundio.

AGRAÏMENTS

Seria mentida dir que aquesta tesi és només mèrit meu ja que molta gent m'ha recolzat acadèmicament i personalment ajudant-me a superar totes les dificultats que m'he anat trobant.

Primerament agrair al Xavier Estivill, per obrir-me les portes del seu laboratori i a la Mariona Bustamante per donar-me l'oportunitat de començar aquest projecte al seu costat. Gràcies Mariona per tant suport i aprenentatge durant aquests quatre anys perquè sense tu no hauria arribat fins aquí. Agrair també a l'Eulàlia Martí per tutoritzar aquesta tesi durant els últims dos anys i per estar al meu costat en alguns moments difícils. I també a tots els membres del projecte HELIX i del consorci PACE per l'ajuda rebuda.

I al laboratori, ai el nostre laboratori... vaig arribar trobant-me tres magnífiques amfitriones: l'Anna, que de seguida es va convertir en confident i amiga, gràcies per tantes converses i consells; la Marta, amb qui vaig trobar uns mateixos ideals, gràcies per tants riures i per haver compartit tants moments de tensió seguint l'actualitat; i la Tere, mi Murciana, siempre animando al laboratorio, gracias por esas risas y momentos de complicidad. I als pre-docs: al Joan, com costava que baixessis a dinar i lo bé que t'ho passaves quan ho feies; al Dani, que tot i compartir poc temps al laboratori hem compartit molt fora, gràcies per tanta ajuda en tants moments difícils; i a la Laura, l'altra supervivent, fins al final superant els obstacles, gràcies per estar allí, fins al final. I a la Georgia, perquè tu mai fallaves quan venies, dinar i cafè, gràcies per totes les converses de cafè. I a la Vero i a la Kelly,

sovint absents però sempre allà quan se les necessitava. I als estudiants que heu anat passant, no posaré noms perquè no voldria deixar-me ningú però cadascun de vosaltres vau aportar frescor al laboratori i tots vam aprendre mútuament.

Agrair també als meus amics, em sento molt afortunada de tenir molt bons amics que estan al meu costat quan les coses van bé però també quan no hi van tant. Els amics castellers, els últims en arribar a la meva vida però que s'han convertit en gent molt important, l'Oriol, la Laia, l'Edu, el Xavi, l'Anna i tota la família borinota que em va acollir quan em sentia perduda a una ciutat tant gran. Gràcies supervivents per tant en tan poc temps, pels dies de Diada, les festes i les confidències. Als Hobbits, Noelia, Bea i Joan, costa que coincidim els 4 però seguim fent els nostres sopars, encara que sigui cada dos anys. Als de Tarragona, el meu gran refugi, als que us trobo tant a faltar, Tania i Guillem, que us veig per Nadal i en tinc prou per saber que res ha canviat; Marina, que just ara també has marxat a l'altra punta de món però hem compartit massa i sé que encara que lluny, estaràs amb mi; A les MaNuHé, la Núria, companya de grans dies i nits compartides i que sé, que tot i veure'ns cada cop menys estàs allà i que això no canviarà, i l'Helena, no sé què faria si no pogués trucar-te, sense els teus consells sobre qualsevol cosa; Alba i Marina, que potser hem compartit menys, però els dies i les nits Tarragonines no serien el mateix si no hi fóssiu valtros. A les Vila-rodonines, la Laura i la Serra, les de sempre, les de tota la vida, sobretot a tu, Serra, que hi has sigut sempre en els millors i els pitjors moments, tota la vida al meu costat sent una part imprescindible de

mi i sé que això, per temps que passi, mai canviarà. Gràcies a tots per ser-hi.

I a la família, als tiets, perquè sempre ajuden els cafès al Miami per desconnectar una estoneta; al Roger, cosí, germà i amic, des que vas néixer al meu costat, recolzant-me; a la família política, que en l'últim any també ha sigut important, tants berenars de desconnexió; i sobretot a valtros, pare i mare, perquè està clar que sense valtros no hauria arribat on sóc ara. Gràcies per tants esforços perquè pogués dedicar-me al que volia, gràcies per ensenyar-me tant a valorar les coses i gràcies per estar allà quan les coses no van bé.

I a tu, Joan, que has estat allà cada dia d'aquest últim any tant dur de tesi, que has estat allà animant-me quan feia falta i aguantant la muntanya russa que ha sigut la meva vida. Gràcies per donar-mho tot i per fer que aquest últim any fos una mica menys difícil.

ABSTRACT

Maternal smoking during pregnancy (MSDP) and postnatal secondhand smoking (SHS) are ongoing public health concerns that are associated with adverse child health outcomes, but not much is known about the underlying molecular mechanisms.

We investigated the association between MSDP and placental DNA methylation, and its link with reproductive outcomes through a meta-analysis. We identified 1224 differentially methylated CpGs in placenta, which were enriched for pathways related to inflammation, growth factors and vascularization. Moreover, the methylation of many of these CpGs was associated with gestational age and birth size.

We also investigated the association between MSDP and postnatal SHS, and different molecular layers in children: blood DNA methylation and transcription, plasma proteins, and serum and urinary metabolites. *In utero* exposure, was only associated with child DNA methylation, confirming a persistent MSDP-related signature on the blood epigenome. However, this imprint was not mirrored in the child transcriptome. In contrast, postnatal SHS was related to protein and metabolite levels, which are more dynamic, and likely reflecting short-term exposures. Of note, the MSDP-associated methylome signature was tissue-specific.

RESUM

Fumar durant l'embaràs i l'exposició a tabac passiu de manera postnatal són preocupacions en la salut pública que estan associades a conseqüències adverses en la salut infantil, però poc se sap sobre els mecanismes moleculars.

Hem investigat l'associació entre fumar durant l'embaràs i la metilació de l'ADN placentari, i la seva relació amb les conseqüències reproductives a través d'un meta-anàlisi. Hem identificat 1224 CpGs diferencialment metilats a placenta, enriquits en vies relacionades amb inflamació, factors de creixement i vascularització. A més a més, la metilació de molts CpGs està associada a l'edat gestacional i les mides al naixement.

Hem investigat també l'associació entre fumar durant l'embaràs i l'exposició a tabac passiu de manera postnatal, i diferents capes moleculars en nens: metilació de l'ADN i transcripció en sang, proteïnes en plasma, metabòlits en sèrum i orina. L'exposició uterina, estava només associada amb la metilació de l'ADN del nen, confirmant un efecte persistent del patró de l'epigenoma en sang. Tot i això, aquesta empremta no es veu reflectida en el transcriptoma. En canvi, l'exposició a tabac passiu estava relacionada amb proteïnes i metabòlits, que són més dinàmics, i reflecteixen efectes a més curt termini. Hem observat també que el patró de metilació és específic de teixit.

PREFACE

A high proportion of the population has been or is in contact to tobacco smoking in an active or a passive way despite being alert of the adverse health effects.

The prenatal period and the early life are critical periods, and environmental impacts during these years can result in later health consequences. The adverse health effects of *in utero* exposure to tobacco smoking have been widely studied and reported. In spite of this, the underlying molecular mechanisms are only partially understood. Also, not much is known about the effects of secondhand smoking in children.

This work wants to contribute to the understanding of the biological mechanisms underlying the effect of prenatal and early life exposure to tobacco smoking on child health, and to identify molecular features that could be used in the development of novel biomarkers of exposure to tobacco smoking.

INDEX

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 Tobacco smoking

Cigarette smoke is a toxic and carcinogenic dynamic complex mixture containing an estimate of 5000 chemicals, 90 of which are identified as harmful according to US Food and Drug Administration (FDA),[1] and its adverse effects on health have been established for many years.[2,3] Some compounds occur naturally in tobacco (e.g. nicotine), others are added during the manufacturing processes (e.g. ammonia), but most of them are products resulting from combustion.[4] Furthermore, some other factors can change levels of mainstream and side stream compounds (e.g. concentration of smoke, specific properties of tobacco or cigarette design and manufacture) converting it in a changeable mixture.[5]

The main harmful compounds of cigarette smoke are shown in Figure 1. Nicotine, which is an alkaloid that can pass through biological membranes being the nervous system its main target, is the fundamental cause of addiction due to its similarity to the brain neurotransmitter acetylcholine.[5,6] Ammonia can cause irritation of airways and other breathing problems.[4] N-nitrosamines, a group of chemical compounds containing a nitroso group attached to an amine nitrogen, are carcinogenic.[7] Polycyclic aromatic hydrocarbons (PAHs), a large group of more than 500 organic compounds with two or more aromatic rings, result from the burning of biological material, and are very hazardous to human health being some of them carcinogenic. Volatile compounds can determine the overall toxicity of tobacco smoking and include carbon monoxide (CO) and

aldehydes (acetaldehyde, formaldehyde or acrolein). Finally, some toxic metals such as lead, cadmium and arsenic, whose levels can be determined by tobacco growing conditions, and other harmful compounds such as aromatic and heterocyclic amines, are also found in tobacco smoking.[5]

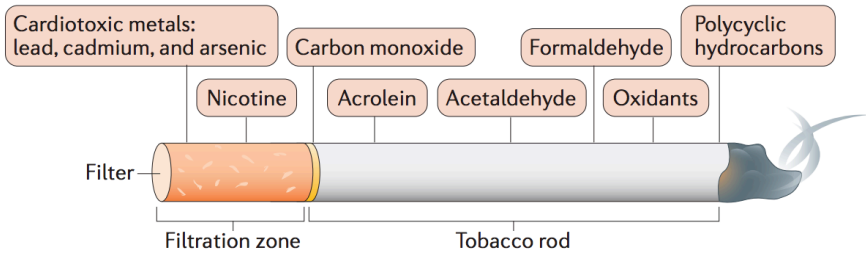

Figure 1. Main cigarette smoking compounds.[8]

The exposure to tobacco smoking can be direct through active smoking or indirect receiving the exposure of other people's smoke, which is called passive smoking. During pregnancy, unborn children can be also exposed to tobacco smoking through active or passive maternal smoking.

1.1.1 Current tobacco smoking: prevalence and health outcomes

Despite the public health awareness and the efforts to promote smoking cessation, the current adult smoking prevalence is still high. From 2007 to 2015, there was a declining from 23.5% to 20.7% of smokers, but due to the population growth, the number of smokers

remained in 1.1 billion people.[9] As shown in Figure 2, Asian countries have higher prevalence while Australia and America show lower percentage of smokers.[10] In Europe, rates are still high in most of the countries.

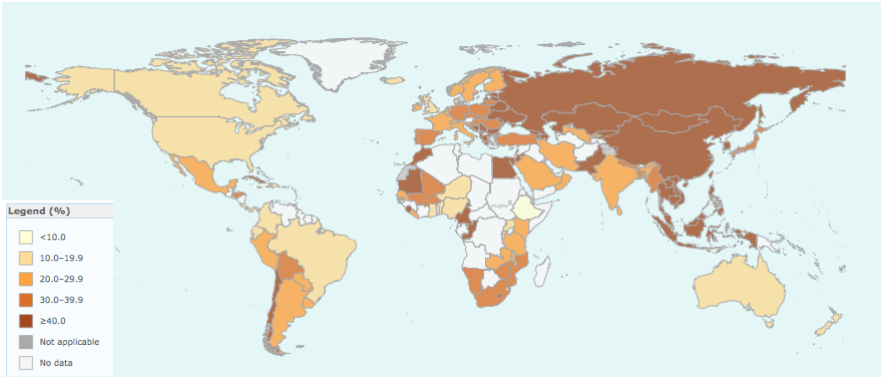

Figure 2. World map showing age standardized current smoking prevalence among adults aged >15 in 2015.[10]

The primary target organ of tobacco smoking is lung, causing injury by a direct chemical exposure to cigarette smoking. It also produces systemic effects resulting of indirect consequences of the exposure (e.g. oxidative stress, inflammation, endothelial dysfunctions and coagulation).[11] Nearly six million deaths per year are caused worldwide by smoking[3] due to an increased risk of developing common diseases: different types of cancer (e.g. bladder, cervical, esophageal, kidney, laryngeal, leukemia, lung, oral, pancreatic and stomach cancers); cardiovascular diseases (e.g. abdominal aortic aneurysm, atherosclerosis, cerebrovascular disease and coronary heart disease); respiratory diseases (e.g. chronic obstructive pulmonary disease, pneumonia); reproductive effects (e.g. fetal death and stillbirths, lower fertility, lower birth weight and pregnancy

complications); and other (e.g. elevated blood pressure, cataract, diminished health status and morbidity, hip fractures, low bone density and peptic ulcer disease).[12]

1.1.2 Secondhand tobacco smoking: prevalence and health outcomes

Secondhand smoking (SHS), passive smoking or environmental tobacco smoking (ETS) is the involuntary exposure to a dynamic mixture of air and the smoke from burning tip of cigarette whose characteristics and concentration changes depending on time and distance from combustion. It is one of the most widespread exposures in the indoor air pollution. According to the World Health Organization (WHO), more than a third of the population is regularly exposed to tobacco smoke (40% of children, 33% of adult male non-smokers and 35% of adult female non-smokers) being Eastern Europe, the Western Pacific and South-East Asia the most affected areas.[10-13] Regardless the bans, in Europe (2014), it has been estimated that around 60% of adult males, adult females and children are still exposed to SHS (Figure 3).[10,14]

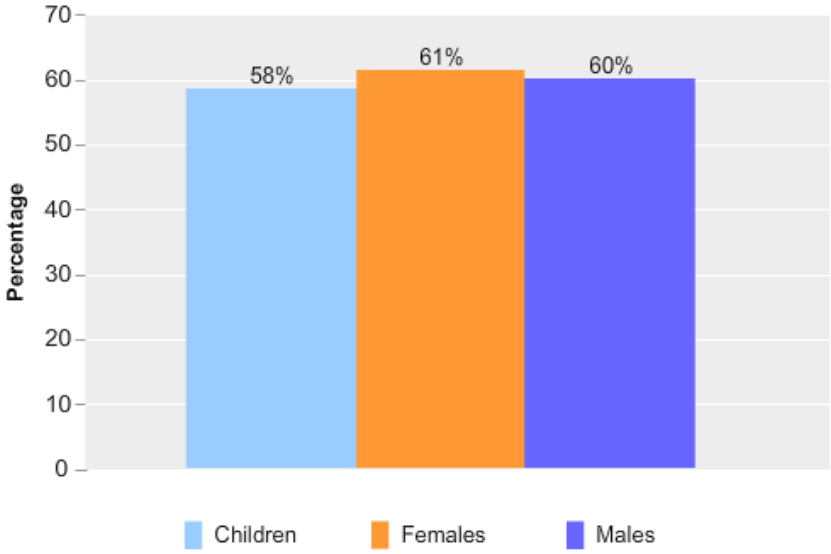

Figure 3. Percentage of children, females and males exposed to SHS in Europe in 2014[10,14]

Exposure to SHS is associated with lung cancer, cardiovascular diseases and respiratory diseases. Children and newborns are especially susceptible to its harms and they have an increased risk to long term consequences as their respiratory and immune systems are not completely developed.[15] Exposure to SHS in children has been associated with an increased risk for many diseases such as otitis,[16] leukemia,[17] obesity[18] and respiratory diseases (e.g. lower respiratory-tract infections, increased risk of wheezing, asthma, airway hyperresponsiveness, impaired lung function, bronchitis or pneumonia).[19–23] Furthermore, exposure to SHS has been also linked to behavioral and intellectual impairments[24] and to a greater risk of sudden infant death syndrome (SIDS).[20]

1.1.3 Maternal smoking during pregnancy: prevalence and health outcomes

Active maternal smoking during pregnancy (MSDP) represents one of the most important avoidable risk factor and its short and long-term adverse effects in offspring have been widely studied. [25] Both, duration (any vs sustained smoking) and intensity of maternal smoking during pregnancy (number of cigarettes per day), are important factors in determining future health consequences.[26] Less is known about the effects of maternal SHS and the results are not as consistent as active smoking.[27]

Despite public health efforts to promote smoking cessation, 8.4% of women smoked during pregnancy in The United States (the percentage ranges from 1.8% in California to 27.1% in Virginia).[28] In European countries, it has been estimated that maternal smoking during pregnancy ranges between 4.2% in Iceland and 18.9% in Croatia. The prevalence not only depends on the region but also on the age and socioeconomic status.[29] In Figure 4 it is shown the prevalence of daily smoking in women and pregnant women by region.[30]

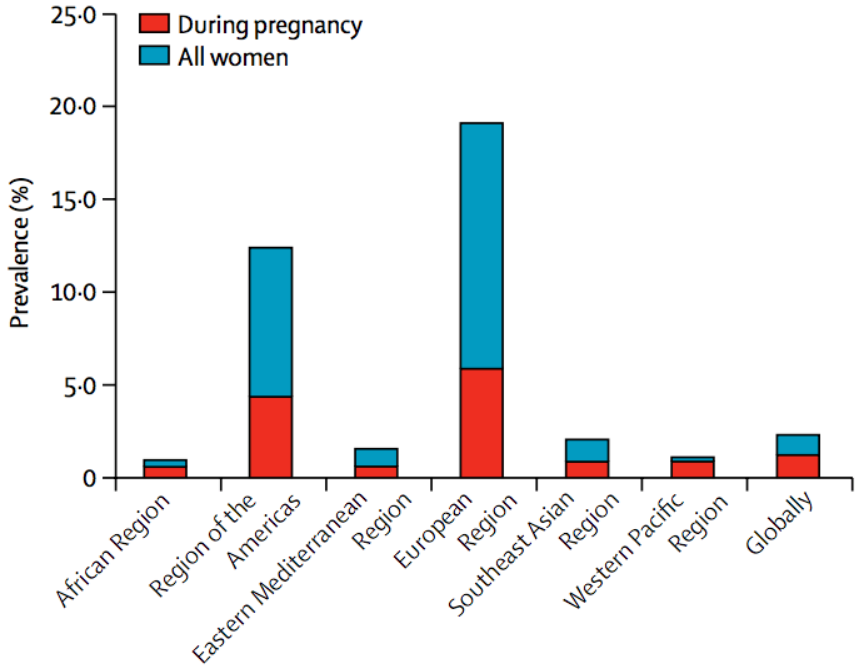

Figure 4. Prevalence of daily smoking in women and pregnant women by region and globally.[30]

The sections below describe the effects of MSDP on pregnancy complications and offspring health.

1.1.3.1 Prenatal exposure to tobacco smoking and maternal adaptations to pregnancy

Different physiological and anatomical alterations occur in different organs and systems of pregnant women to favor the correct fetal growth and development,[31] and exposure to tobacco smoking can affect them.

In first place, active smoking can negatively affect the ovule fertilization making more difficult the start of pregnancy, and it increases the risk of abortion.[25]

The maternal immune system is adapted to maintain pregnancy and prevent fetal rejection. There is a shift from Th1 response (i.e. inflammatory response) to Th2 response to favor maternal tolerance. MSDP can increase the production of inflammatory cytokines, chemokines and growth factors producing a shift to a Th1 response, non-favoring maternal tolerance. MSDP is also associated with an alteration of NK cell cytotoxic functions, which makes the mother more vulnerable to infections.[32] Overall, MSDP decreases the maternal immune response.[33]

The maternal hematological system also suffers some changes. Plasma volume increases progressively during pregnancy and subsequently, there is a fall in hemoglobin concentration, hematocrit (i.e. volume of erythrocytes) and red blood cell counts. Moreover, there is a change in the coagulation system producing a physiological hypercoagulable state, which predisposes the pregnant woman to venous thrombosis.[31] MSDP enhances this platelet aggregation, reduces the dispensability of blood vessel walls and induces prothrombotic and proinflammatory state.

Pregnancy is a diabetogenic state and changes in glucose metabolism are necessary to a correct shunting of glucose to the fetus to ensure

its development while maintaining maternal nutrition.[31] The diabetogenic state is a result of increasing secretion of diabetogenic hormones. MSDP decreases the levels of certain of these hormones and growth factors that play an important role in fetal growth.[34] There is also an increase in total serum cholesterol and triglycerides to provide energy to the mother while glucose is spread to the fetus. MSDP increases the serum levels of cholesterol and triglycerides but with a tendency towards an adverse lipoprotein profile, decreasing the concentration of high-density lipoprotein cholesterol and increasing low-density lipoprotein cholesterol.[33]

Changes in maternal endocrine systems includes increased levels of thyroid and corticoid hormones to ensure a normal development of brain and nervous system.[31] MSDP alters the levels of important hormones for a correct fetal growth.

1.1.3.2 Prenatal exposure to tobacco smoking and placental function

Placental abruption and placenta previa are negative health consequences associated with MSDP and they are related to fetal and maternal morbidity and mortality. Placenta previa occurs when placenta implants in the poorly vascularized lower uterine segment, which lead to a lower uteroplacental perfusion and subsequently to a fetal growth restriction.[35] Abruption occurs when placenta partially or completely separates from the uterus too early. MSDP can cause decidual necrosis at the periphery of the placenta, microinfarcts,

fibrinoid changes in the placenta and hypo vascular and atrophic villi leading to placental abruption and a reduction of uteroplacental-fetal blood flow.[22] Placentas from smoker mothers also have reduced nutritive and excretory functions.[36]

1.1.3.3 Prenatal exposure to tobacco smoking and preeclampsia

Preeclampsia is a systemic syndrome of pregnancy characterized by an increase of maternal blood pressure and proteinuria affecting both mother and fetus.[37,38] Paradoxically to all the adverse effects of smoking during pregnancy, a meta-analysis of several studies[39–55] have associated MSDP with a 33% decreased risk of preeclampsia.[37] However, the biological mechanisms of this inverse association are not completely understood. The most likely mechanism is the reduction in smoker mothers of soluble vascular endothelial growth factor receptor-1 (VEGFR1) and soluble endoglobin (sEng) levels, which are important antiangiogenic factors in preeclampsia pathogenesis.[37,38] It has been observed that an imbalance of endogenous angiogenic factors plays an important role in the pathogenesis producing endothelial dysfunction[56], which is a common mechanism involved in several cardiovascular diseases and also with higher blood pressure[57,58] and proteinuria.[59,60]

1.1.3.4 Prenatal exposure to tobacco smoking and reproductive outcomes

The adverse health effects of MSDP at birth have been largely studied for years, with the first association reported in 1957. MSDP increases the risk of premature birth, prior to 37 weeks of gestation, by more than 25%,[61] and children of smoker mothers weight between 150 g and 200 g less than those of non-smoker mothers.[62] Some studies have suggested that the reduction of birth weight associated with MSDP is predominately mediated by fetal growth restriction, rather than by shorter pregnancy duration.[63]

A dose-dependent relationship has been observed for both, preterm birth and birth weight. SHS of pregnant women, by parent's smoking in pregnancy or other, can also derive to a lower birth weight, although the effect is lower than in active smokers (Figure 5).

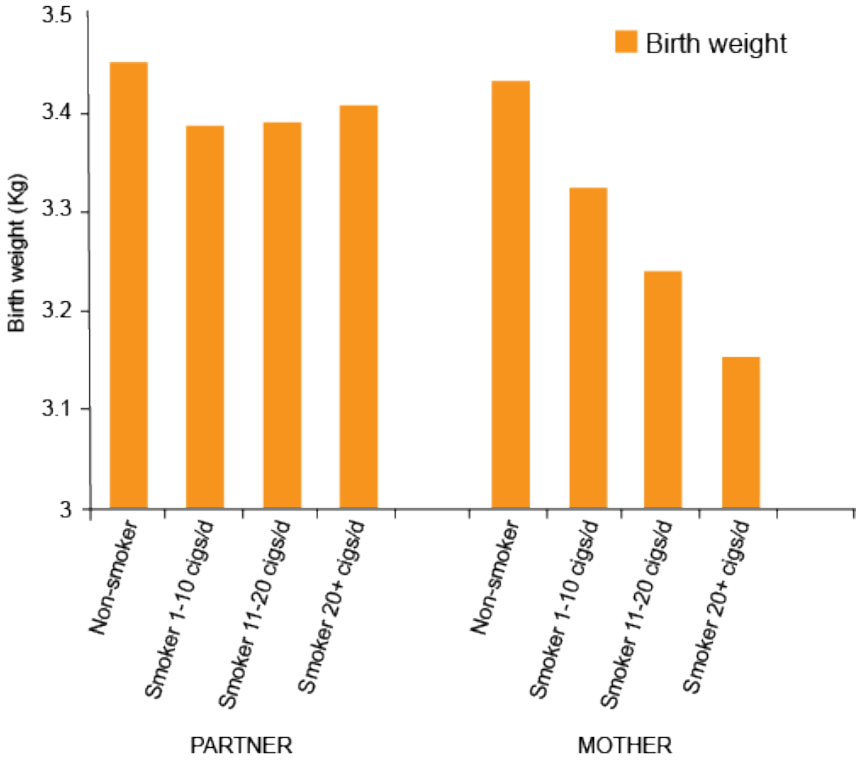

Figure 5. Estimated effects of maternal smoking at any point during pregnancy or exposure to SHS (by partner's smoking status) on birth weight (BW).[62]

The mechanisms that link MSDP with adverse reproductive outcomes are partially known.[63] These are some of the several possible mechanisms: i) reduction of maternal appetite caused by smoking resulting in poor fetal nutrition; ii) vasoconstriction caused by smoking which would decrease fetal blood supply (e.g. nicotine reduces the uterine blood flow);[62] iii) direct toxic effect of smoke compounds on fetal metabolism (e.g. cadmium accumulates in the placenta and inhibits the activity of an enzyme associated with intrauterine growth); iv) increased susceptibility to vaginal infections and attenuated immune response; v) increased risk of membrane

rupture; and vi) hypoxia of the fetus caused by elevated CO levels (Figure 6).[22]

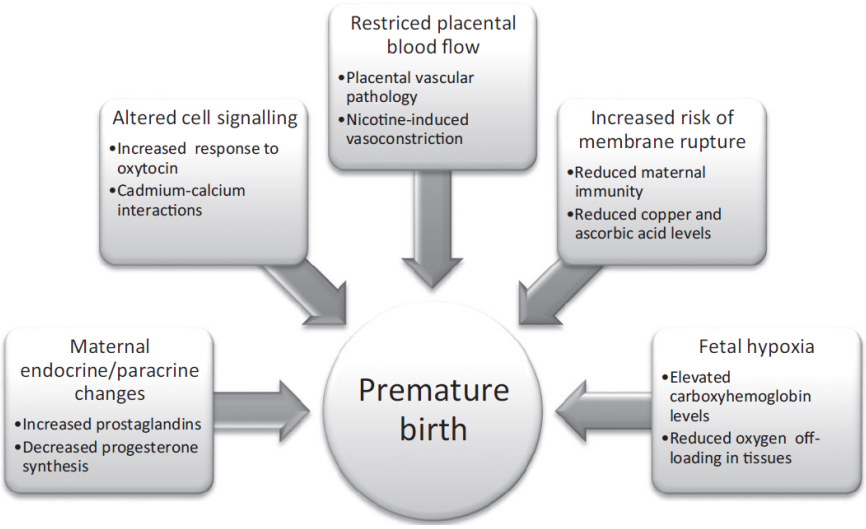

Figure 6. Mechanisms through which MSDP may affect gestational age at delivery.[63]

Both prematurity and lower birth weight are the leading cause of newborn death, with more than 1 million children dying form preterm birth-related problems each year worldwide.[64] It has been described an increased risk of stillbirth and neonatal death during the first year of life in smoker mother's offspring.[62,65]

1.1.3.5 Prenatal exposure to tobacco smoking and later child health consequences

Besides the effects of MSDP on reproductive outcomes, MSDP has also been extensively related to later heath consequences in the

offspring, including intellectual and behavioral problems, reduced lung function and asthma, obesity and cardio-metabolic problems.[23]

Some of these later effects might be mediated through fetal growth restriction or shorter gestation, but others can be caused by the direct effect of tobacco compounds in organ development. For instance, nicotine, which can cross placenta barrier,[66] can directly affect lung and brain development having negative health consequences later in childhood (e.g. respiratory, cognitive, emotional and behavioral problems).[62,67]

Although MSDP results in lower birth weight, this changes completely in childhood with an increased risk of developing overweight and obesity.[68] Postnatal catch-up growth is one of the proposed mechanisms.[23] MSDP is also associated with higher systolic or diastolic blood pressure with an increased risk of cardiovascular diseases.[23] Persistent endothelial dysfunction, changes in renal structures and function, and alteration in perivascular adipose tissue might explain this association.[23]

MSDP is shown to produce an increased risk of asthma and wheeze in children and adolescents, with the strongest effect in children aged ≤2 years.[69] It has been proposed that in utero tobacco exposure may enhance the allergic inflammatory response. On the other hand, different studies have reported that *in utero* tobacco exposure has been associated with a decrease in lung function,[23] which seems to persist into adolescence, independently of asthma.[70]

Finally, MSDP has been associated with offspring behavioral problems and psychiatric disorders, particularly conduct disorder, attention deficits, hyperactivity, substance abuse, and depression.[71] Moreover, an inverse relationship between MSDSP and offspring intelligence, cognitive ability and school performance has been described in numerous studies.[71] The mechanisms have not been elucidated, but it is thought that hypoxia and malnutrition could be relevant. Also, nicotine could interfere in the normal function of acetylcholine in fetal brain development.

1.2 The internal exposome

Although many studies have evidenced an association between tobacco smoking and health outcomes, the underlying biological mechanisms of these findings are not completely understood, especially for the prenatal exposure to tobacco smoking.

The set of molecular factors that can change in response to environmental factors is considered the internal exposome, included in the global exposome concept. The exposome concept was introduced by first time in 2005 by Dr. Christopher Paul Wild, as a need to match it with the genome developing methods to assess the individual's environmental exposures with the same precision as the ones existing for the individual's genome.[72]

The exposome represents the entirety of environmental exposure from conception to death, and its highly variable and dynamic behavior make it extremely challenging to measure.[72] The exposome encompasses both external, which is further categorized as general and specific, and internal exposures. External exposome refers to the socioeconomic environment (general) as well as the personal exposures of each individual (specific).[73,74] Finally, the internal exposome reflects the biological effects of the external exposure and the biological responses to them which are needed to maintain homeostasis. A schematic explanation of the exposome concept is shown in Figure 7.

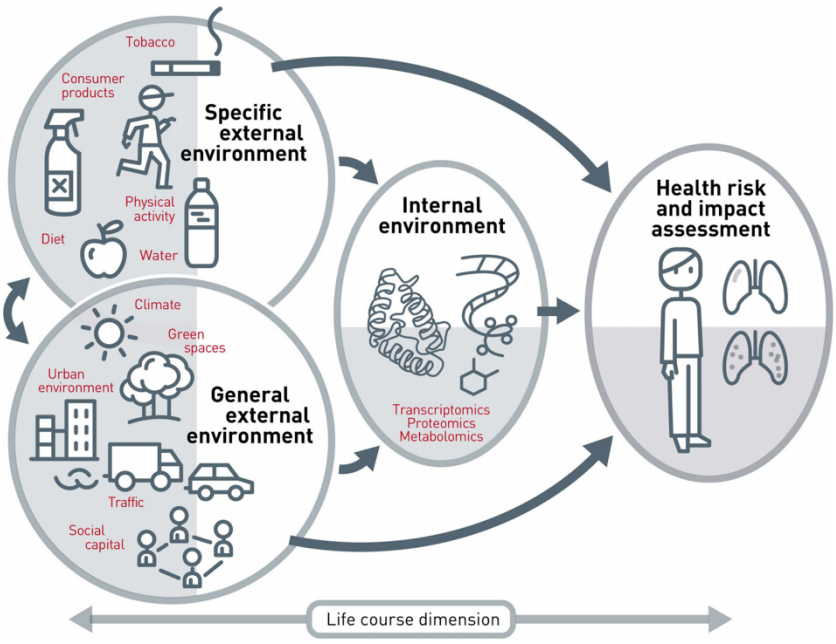

Figure 7. The exposome concept: specific and general external environment, internal environment and its influence in health.[75]

The internal exposome is dynamic over time because it controls developmental and physiological processes, and because it responds to environmental factors, which are changeable.

The study of the internal exposome can provide the following information:

1. A better understanding of the molecular mechanisms through which environmental exposures end up affecting health.[76]
2. The development of novel and more accurate biomarkers of exposure that improve the existing ones.

1.2.1 Molecular layers of the internal exposome

The internal exposome includes different molecular layers: the epigenome, transcriptome (coding and non-coding RNA), proteome and metabolome[73]. Epigenetic marks are the most widely investigated factors.

Epigenetics is the study of any process that modifies gene activity without changing the underlying DNA sequence.[77] These modifications are produced by changing the chromatin architecture and accessibility. They play an important role regulating gene expression, which in turn can affect protein and metabolite levels.[78] Several epigenetic marks are known: DNA methylation, regulatory RNA and histone modifications (Figure 8).

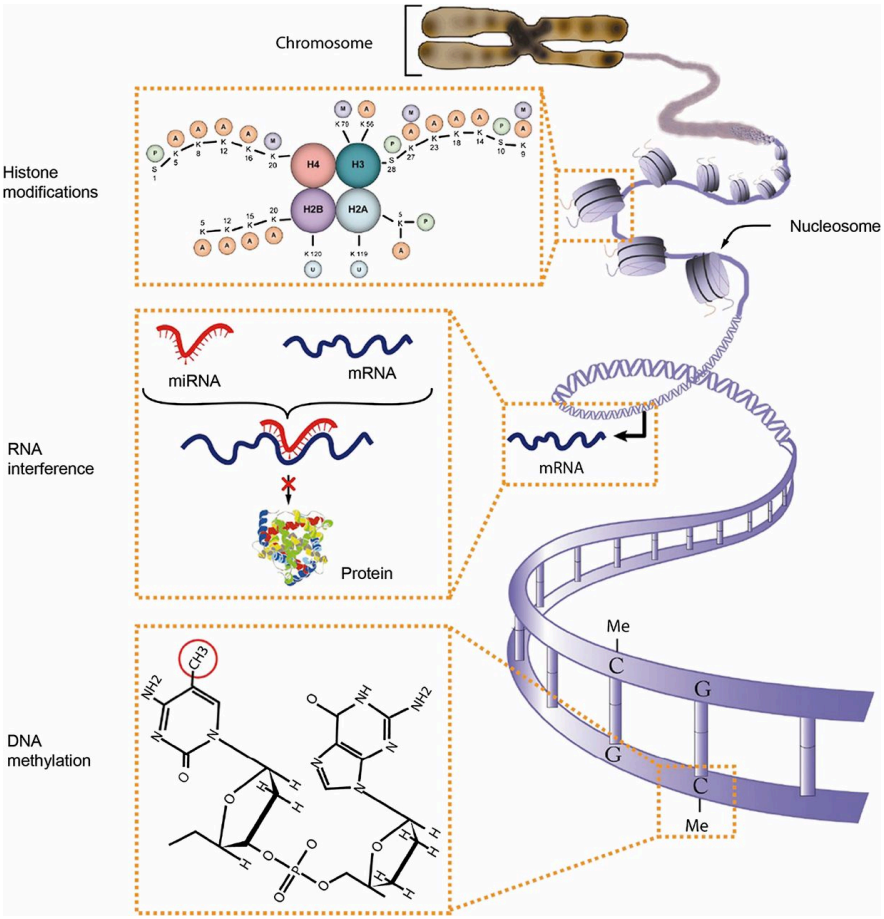

Figure 8. Scheme of epigenetic regulation mechanisms: DNA methylation, histone modifications and RNA interference.[79]

Chromatin is composed by histones (complex of proteins), which can be chemically modified. Histone modifications (acetylation, methylation, phosphorylation, ubiquitylation and sumoylation) occur on lysine, arginine and serine residues within histone tails. The combination of these modifications can open the chromatin, which make it more accessible and subsequently activate gene expression,

or they can fold chromatin, which makes less accessible and shut expression down.[77,78,80]

DNA methylation is the most widely studied modification as it can be easily assessed with current methods. It consists in the addition or removal of a methyl group (CH_3) to the 5-prime position of a cytosine in the context of a CpG. These cytosine residues can be located in cytosine/guanine-rich stretches of DNA called "CpG islands", which are often found in regulatory regions of the genome.[77,78] Generally, when a CpG is methylated, the gene is silenced and, subsequently, it is not expressed. It is estimated that, in mammalian genomes, about 80% of CpG dinucleotides are methylated but CpGs dinucleotides located in islands are usually unmethylated.[81]

Non-coding RNAs are RNA molecules that do not code for any protein, but have a regulatory role, so they are considered epigenetic mechanisms. Non-coding RNAs are divided according to their size and mechanisms of action. miRNAs are single strand RNA of 21-24 nucleotides which silence genes when they base-pair with complementary sequences of the target mRNA in 3'UTR by translational repression or mRNA cleavage.[78] One miRNA can regulate the expression of one or more genes. The formation of mature miRNAs starts with the transcription of the primary miRNA (pri-miRNA) containing poly-A tails. This primary miRNA then forms a hairpin and is digested by the dsRNA-RNA-specific ribonuclease Drosha forming the precursor miRNA (pre-miRNA) which is 70-75 nucleotides. The precursor miRNA associated with

the RNA-induced silencing complex (RISC) is transported from the nucleus to the cytoplasm where it is cleaved by Dicer (an RNase III superfamily member) resulting in a double strand. One strand remains the mature miRNA while the other one is degraded (Figure 9).[81]

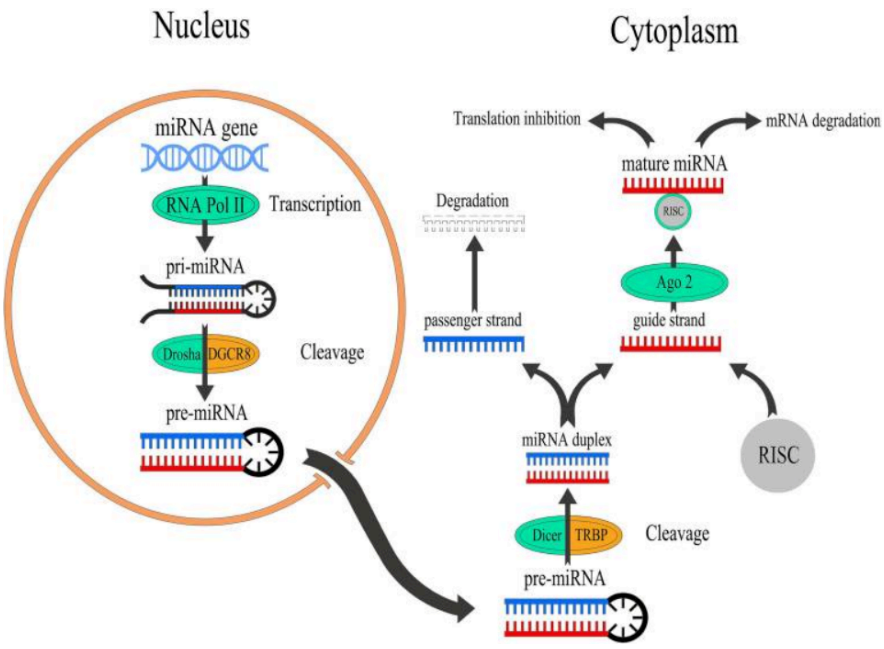

Figure 9. Overview of miRNA biogenesis.[82]

Epigenetic modifications interact among them forming a multiple layers of gene regulation.[78] The combination of different epigenetic marks in a particular region of the genome gives place to a genomic regulatory element (e.g. transcription start sites, enhancers, active genes, repressed genes, exons or heterochromatin). These regulatory segments of the chromatin are named "chromatin states" and have been described for different tissues in the context of the NIH

Roadmap Epigenomics Consortium,[83,84] a public resource of human epigenomic data, and the Encyclopedia of DNA Elements (ENCODE) project.[8586]

Finally, the transcriptome is defined as the full range of RNA molecules expressed in a specific cell or tissue, at a particular developmental stage and under a certain physiological or pathological condition. It comprises coding (messenger) and non-coding (ribosomal, transfer, small nuclear, small interfering, micro and long-non-coding) RNAs, which represents a small percentage of the genome transcribed to total RNA (less than 5% in humans).[87,88] The proteome and the metabolome are the global collection of proteins and metabolites expressed in a cell or tissue in a particular time. Proteins and metabolites are highly dynamic and its changeability makes difficult to reflect the entire picture and repeated measures are needed to capture short-short term responses.[89]

1.2.2 Embryogenesis and fetal development

Epigenetic regulation is involved in many biological processes, one of them embryogenesis and fetal development. Fetal development is a complex process which needs carefully regulated cell proliferation, the formation of distinct cell lineages that will adopt unique cell functions and coordinated interactions between cells to produce all the distinct tissues.[90]

After fertilization, the zygote experiences a series of cell divisions to form a mass of cells which is called morula. After compaction of the morula, the first differentiation event occurs to form the blastocyst when it has passed from the oviduct into the uterus where the blastocyst is implanted in the uterine wall. In this stage, a ring of trophoblast cells, which will form the placenta(see below, section 1.2.3.2), surrounds the inner cell mass, which will form the embryo (Figure 5).[91] The embryo is located in the amniotic cavity, which is separated from the chorionic plate of the placenta for a thin membrane called amniotic membrane. This amniotic membrane is formed from trophoblastic cells and it has multiple metabolic functions and also protect the fetus.[91]

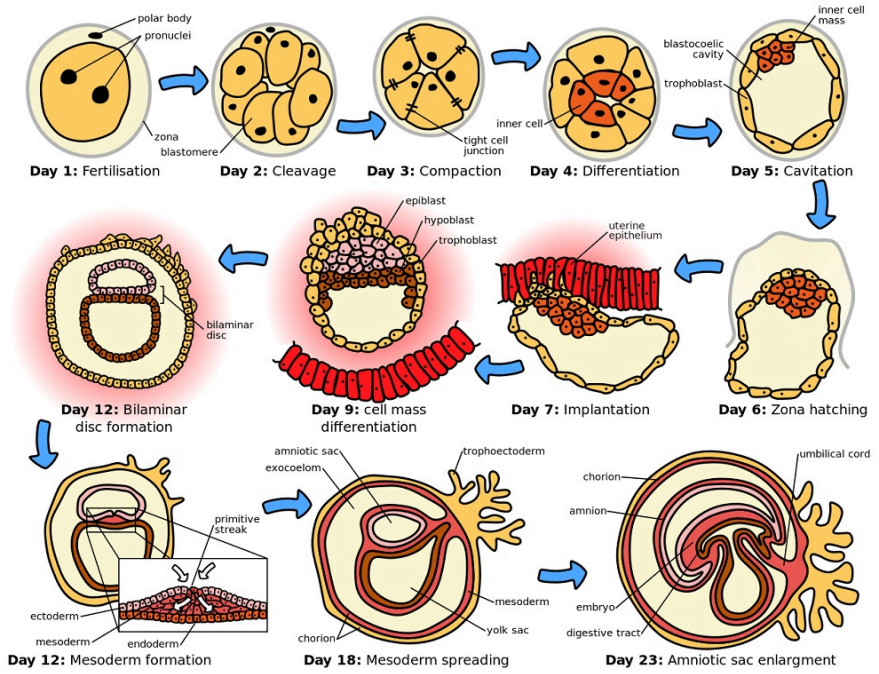

Figure 10. Scheme of embrionogenesis[92]

During the third week the embryo undergoes the process of gastrulation which stablishes the three definitive germ layers (endoderm, mesoderm and ectoderm) that will differentiate into the different organs. The ectoderm will give raise to the nervous system and the epidermis, the mesoderm to the muscle cells, connective tissue and some organs (e.g. heart, kidneys, uterus, ovary or testis) and the endoderm to the gut and many internal organs (e.g. stomach, intestines, respiratory tract, liver, or pancreas). Moreover, mesoderm and trophoblast tissue will form the umbilical cord.[93]

From the implantation and until the eighth week (embryonic phase), different organ systems differentiate at different rates. During this phase the major internal and external organs are formed although with only few cells. After eighth week starts the fetal period where there are two different processes: a process of rapid growth and a process of tissue and organ differentiation in which cells are specialized for different functions. The total process lasts 40 weeks.

All these processes are accompanied with several waves of demethylation and remethylation. After conception, the paternal genome undergoes active demethylation, and the maternal genome is passively demethylated during subsequent cleavage divisions (Figure 11).[94] After implantation, the embryo undergoes de novo methylation. Imprinted genes escape these waves of demethylation and de novo methylation. During gametogenesis, genome-wide demethylation and de novo methylation occur in the male and female

germ cells accordingly with the sex of the fetus. This is critical for the establishment of genomic imprinting in the next generation.

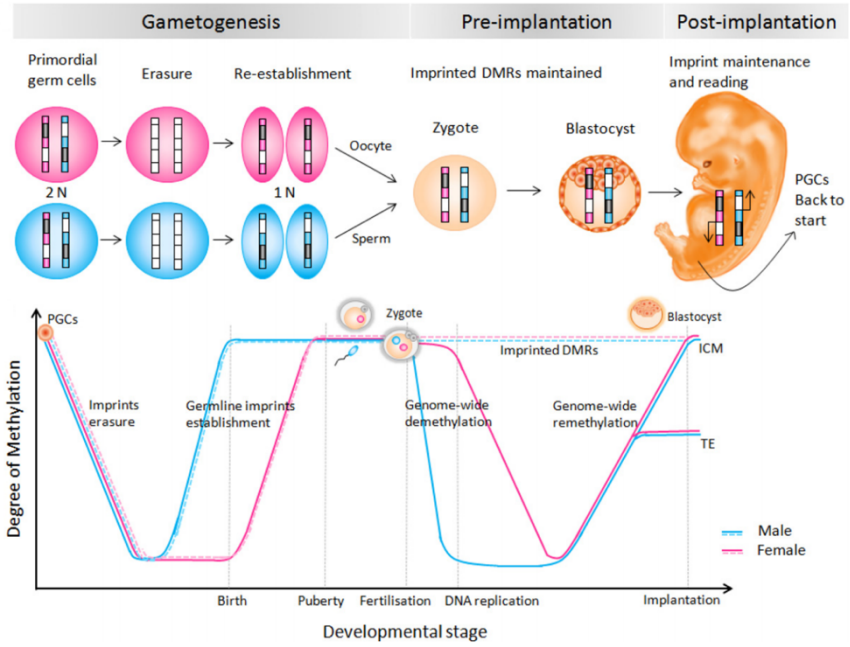

Figure 11. DNA methylation during gametogenesis and fetal development. Primordial germ cells erasure the imprinted methylation reestablishing it during gametogenesis following sex-specific imprints. After the zygote formation, there is a global demethylation in which the imprinted regions are maintained and the genome is re-methylated during implantation.[95]

Imprinting is an epigenetic mechanism of inheritance which consists in the expression of specific genes from only one parent.[96] In these genes, the expression of one of the inherited alleles is suppressed through epigenetic mechanisms. Imprinted loci contain differentially methylated regions (DMRs) providing *cis*-acting regulatory elements which affect expression of surrounding genes. These DMRs are established during gametogenesis (Figure 11). There are several theories to explain the reason of imprinted genes but the most popular is a paternal conflict about nutrient supply and demand. The loss of

these specific imprinting can lead to growth adverse outcomes as well as behavior disorders or other complex traits.

1.2.3 Tissue specificity of the internal exposome

The internal exposome is tissue specific. Tissues exhibit characteristic DNA methylation and transcriptional signatures.[97,98]

In this thesis we have analyzed two tissues: whole blood and placenta. Both tissues have been analyzed because they are accessible. Blood can inform about the immune system; and placenta, a key organ for fetal development, about reproductive outcomes.

1.2.3.1 Blood

Blood is composed by different cell types, which are formed from hematopoietic stem cells (HSCs) through a process called hematopoiesis. Blood includes red blood cells (RBCs), which do not have DNA except in cord blood (nucleated RBCs), platelets and white blood cells (WBCs), which are a mixed population of myeloid and lymphoid cells (Figure 12).[99] Blood cell type proportions change with age and disease status.

It has been observed that each cell type from blood exhibit different methylation patterns,[100,101] which substantially differ between myeloid and lymphoid cells.[102] Given the specific methylation

patterns of blood cell types, and that cell type sorting is not viable in epidemiological designs, deconvolution methods to address potential confounding for cellular composition are needed. They are classified in two types: reference-based and reference-free. More details about these methods can be found in Teschendorff, *et. al.*[103]

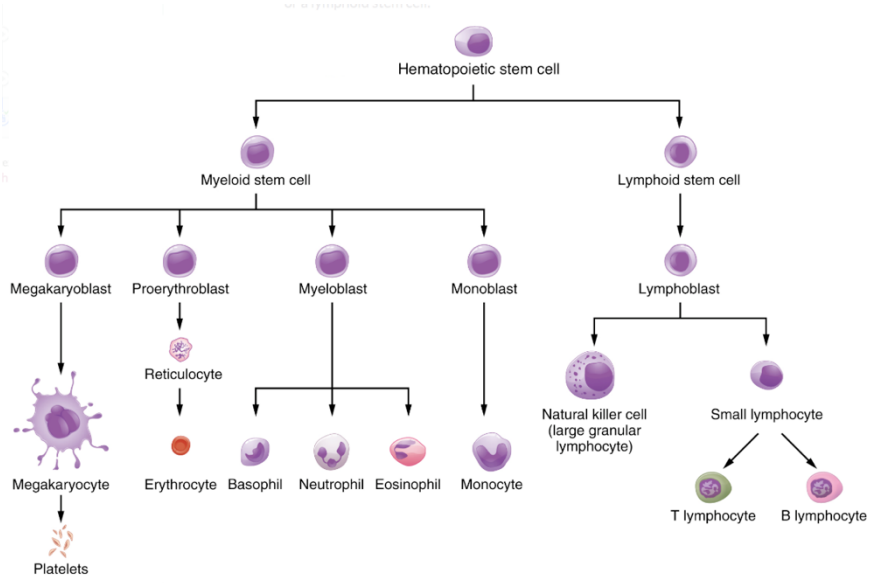

Figure 12. Scheme of blood cell composition.[99]

1.2.3.2 Placenta

Placenta is the first organ to develop during embryogenesis and it is essential for the fetus growth and development supplying it with oxygen, water, and nutrients while removing carbon dioxide and other waste products. Placenta is also an endocrine organ producing

endocrine factors such as estrogens, progesterone, chorionic gonadotrophin, placental lactogen, placental growth hormone, a number of growth factors, cytokines and chemokines. Furthermore, placenta have protective functions from certain xenobiotics and infections.[104]

Trophoblast is the first cell lineage which is differentiated during the firsts five days after conception, between the morula and blastocyst stages, and it encircles the inner cell mass (Figure 10). Later, the trophoblasts give raise to the fetal membranes and the epithelial cover of the placenta whereas the inner cell mass form the umbilical cord and the embryo. From the embryo is formed the extraembryonic mesoderm and it give raise to the placental mesenchyme, originating later the placental fibroblasts, vascular network and resident macrophage population[105,106]. The development of the placenta begins after a week of conception when the blastocyst is attached to the uterine epithelium. The implantation is determined by the trophoblasts that are overlying to the inner cell mass (polar trophoblasts).[107]

During implantation, trophoblasts are merged to form syncytiotrophoblasts which lead the embryo pass through the uterine epithelium into the decidual stroma (i.e. uterine lining, endometrium) and being to the only fetal tissue in contact to maternal cells and fluids. The cytotrophoblasts (remaining mononucleated trophoblasts) are not in contact to maternal tissues but they keep dividing and fusing with syncytiotrophoblast and expanding them. After eight

days from conception, lacunae are formed from the fluid-filling and fusion of the syncytiotrophoblast. The syncytiotrophoblast penetrates to the endometrium and some maternal blood cells go into the lacunae leading to a slow flow of blood. At this stage, the three main parts of the placenta are already determined: the chorionic plate at the fetal side where the umbilical cord is attached; the basal plate at the maternal side adjoined to the endometrium; and the intervillous space, which is the cavity between the two surfaces (a detailed picture of a mature placenta is shown in Figure 13).

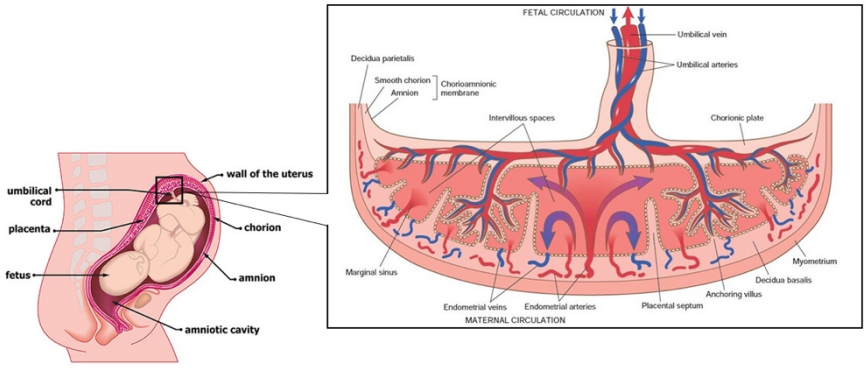

Figure 13. Scheme of anatomy of the placenta. Adapted from[108,109]

Twelve days after conception the implantation is finished and the primary villi, which are syncytiotrohpoblast protrusions, is developed inside the intervillous space. Soon after, the extraembryonic mesodermal cells from chorionic plate go into the intervillous space and the primary villi providing them a mesenchymal core to form the secondary villi. Hematopoietic progenitor cells develop and start to differentiate inside it producing

the first placental blood cells and endothelial cells which generate a new and independent vascular system (tertiary villi).[105] Terminal villi are the final branches of the villous tree, which are richly vascularized by fetal capillary network and the main sites where the maternal-fetal exchange occurs.[106] Utero-placental circulation system allows an exchange of gases and metabolites via diffusion. The mother and the embryo systems are closer but maternal and fetal blood are not in direct contact. A scheme of these process can be found in Figure 14. Therefore placenta is composed by different cell types, being cytotrophoblasts and fibroblasts the main ones, but also endothelial cells, macrophages and mesenchymal cells.[110] Due to the lack of placenta specific references for deconvolution, reference-free methods are applied.

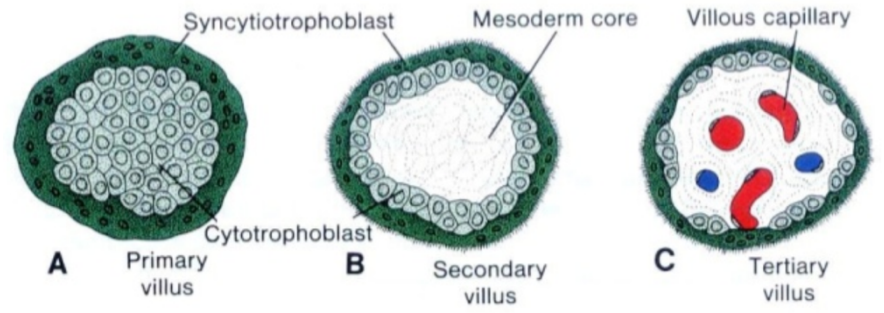

Figure 14. Formation of villi in placenta

Human placenta shows a unique pattern of methylation. It is hypomethylated compared to somatic tissues. Placenta contains partially methylated domains (PMDs), which are large regions (>100 kb) in the genome with a reduced average of DNA methylation levels

and an enrichment in heterochromatin histone modifications.[111] PMDs are stable across the three trimesters of pregnancy and genes located in these domains are generally repressed. Despite PMDs show overall hypomethylation, there is hypermethylation in promoter regions within PMDs.[112]

On the other hand, placenta is enriched in imprinted loci and it is thought that they are important for fetal growth and development, modulation of maternal energy supply and differentiation of trophoblasts cells.[113]

Since placenta is rich in CpGs associated with imprinting control regions and PMDs, which both have intermediate levels of DNA methylation, it shows a trimodal distribution of DNA methylation instead of the bimodal distribution seen in somatic tissues where most of CpGs are, on average, 0% methylated or 100% methylated.

1.2.4 Modulation of the internal exposome

Besides the control of normal developmental processes, the internal exposome also responds to environmental and genetic signals.

Several environmental influences have been related to epigenetic changes such as stress, socioeconomic status, and exposures to various environmental factors including toxicants, nutritional factors, parental body mass index, gestational diabetes and maternal antibiotic use.[114]

In the context of the developmental origins of health and disease (DoHAD) hypothesis, which is based on the study of the impact of early life environmental to the risk of chronic disease from childhood to adulthood and the mechanisms involved, epigenetic modifications have been widely described as the potential bridge between environmental exposures and later health outcomes.[78] According to this hypothesis, the prenatal period and early life represents a major window of susceptibility.

DNA methylation can also be influenced by single nucleotide polymorphisms (SNPs) known as (methylation quantitative trait loci (mQTL)). DNA methylation in these sites is highly allele specific and these changes can reflect different genetic ancestries.

1.3 Tobacco smoking and molecular signatures

Some studies have investigated the link between exposure to tobacco smoking and different molecular layers of the internal exposome. In the next sections, we describe the most important findings from these studies.

1.3.1 DNA methylation

1.3.1.1 Current tobacco smoking

Because of the global health problem of smoking and the suggestion that DNA methylation plays an intermediate role to health outcomes, several studies have analyzed DNA methylation changes in current smokers. Between 2011 and 2015, 17 studies assessed the association of smoking with blood DNA methylation comprising 14 epigenome-wide association studies (EWAs),[115–128] and 3 gene-specific methylation studies.[129–131] Among EWAs studies, 1460 CpG sites were identified to be associated with current smoking, 62 of which were identified in more than three studies.

In 2016, the biggest study was carried out meta-analyzing 15907 blood samples from 16 different cohorts including current smokers, former smokers and never smokers. A total of 18760 CpG sites were identified to be differentially methylated at 5% False Discovery Rate (FDR) (2623 reaching Bonferroni (BN) significance), 16673 of which were not previously reported (1500 reaching BN significance). Top 10 CpGs can be found in Table 1. The results replicated the previous findings in *AHRR, RARA, F2RL3* and *LRRN3*. *AHRR* loci contains the CpGs with the strongest and the most consistent association being cg05575921 the top CpG in different studies.[132] 11627 CpGs out of the 18760 CpGs differentially methylated showed also a dose-response effect. Moreover, several molecular processes

were found to be associated with smoking-related CpGs such as transduction, protein metabolic processes and transcription pathways as well as variants associated with smoking-related phenotypes such as coronary heart disease, cancers, inflammatory diseases, osteoporosis and pulmonary traits. Smoking sensitive CpGs were enriched for island shores, gene bodies, *DNase1* hypersensitivity sites and enhancers suggesting that many of the smoking-associated CpG sites may have regulatory effects.[133]

Table 1. Top ten CpGs associated to current smoking in blood.[133]

CpG	Location	Gene	Effect	P. Value
cg16145216	*1:42385662*	*HIVEP3*	0.030	6.7E-48
cg19406367	*1:66999929*	*SGIP1*	0.018	7.0E-44
cg05603985	*1:2161049*	*SKI*	-0.012	1.8E-43
cg14099685	*11:47546068*	*CUGBP1*	-0.012	1.5E-42
cg12513616	*5:177370977*	*5q35.3*	-0.026	6.1E-41
cg03792876	*16:73243*	*16p13.3*	-0.018	7.2E-38
cg01097768	*5:378854*	*AHRR*	-0.017	6.8E35
cg26856289	*1:24307516*	*SFRS13A*	-0.016	8.6E-35
cg07954423	*9:130741881*	*FAM102A*	-0.013	1.2E-34
cg01940273	2:233284934	*2q37.1*	-0.082	2.0E-34

2568 CpGs showed differentially methylated patterns in former smokers although they had lower effect sizes and the majority of them returned to normal levels within 5 years after cessation. However, 36 CpGs not returned to never-smoker's levels even after 30 years after cessation among which there are *TIAM2, PRRT1. AHRR, F2RL3, GNG12, LRRN3, APBA2, MACROD2* and *PRSS23.*

However, regardless of the abundant literature on the topic, some points still need more investigation, such as the role of ethnicity,

genetic variation, blood cell type composition and the effects on other tissues besides blood.

Most of the studies have been performed in European ancestry populations. Only few have explored other ethnic groups, some of them showing different smoking associated patterns in Asiatics.[120,128,134,135] Methylation status at cg18146737 (*GFI1*), cg05575921 (*AHRR*) and cg12803068 (*MYO1G*) is influenced by *cis-methyl*-quantitative trait loci (meQTLs) suggesting that single nucleotide polymorphisms (SNPs) can also be responsible of the effects of DNA methylation in these genes.[136]

1.3.1.2 Secondhand smoking

Although current smoking and its association with DNA methylation has been widely proved, less is known about the effects of exposure to SHS on DNA methylation. *AHRR* cg05575921 does not show any difference between individuals exposed to SHS and unexposed, except when SHS was categorized in hours.[137] Then, an inverse association between methylation at cg05575921 and the number of hours per week of recent SHS was observed. This study suggests that the effect of SHS on DNA methylation is much weaker than the effect of active smoking. More studies with larger sample sizes and assessing other CpGs are needed to try to elucidate the effects on DNA methylation of exposure to SHS.

1.3.1.3 Maternal tobacco smoking during pregnancy

Several studies have explored the association of MSDP and offspring blood DNA methylation through EWAs analyses in cord blood,[138–141] shortly after delivery,[142] during childhood and adolescence,[140,143] and later in life.[144]

Until now, the biggest study of MSDP and changes in offspring blood DNA methylation was performed by Joubert, *et.al.* in 2016.[141] It consisted of a meta-analysis of 13 cohorts. Cord blood from 6685 newborns (13% of them exposed to MSDP in a sustained way and 25% exposed at some point during pregnancy) and blood from 3187 older children (8% of them exposed in a sustained way and 13% exposed at some point during pregnancy) were analyzed using the HumanMethylation450 (450K) BeadChip. Sustained MSDP was associated with 6073 CpGs in cord blood at 5% FDR, 568 of which remained significant at BN threshold. Any-MSDP was associated to a lower number of CpGs differentially methylated (4653 CpGs) despite of having a major percentage of smokers. The top CpG in all the models was cg05575921 mapped to *AHRR* gene and previously reported in many studies on current smoking.[141] The top 10 CpGs can be seen in Table 2.

Table 2. Top ten CpGs associated to maternal sustained smoking during pregnancy in cord blood (CB) and its persistence after 7 years (7y).[141]

CpG	Location	Gene	Effect (CB)	P. Value (CB)	Effect (7y)	P. Value (7y)
cg05575921	*5:373378*	*AHRR*	-0.064	1.64E-193	-0.013	1.87E-09

cg12803068	*7:45002919*	*MYO1G*	0.070	2.64E-134	0.070	1.31E-22
cg04180046	*7:45002736*	*MYO1G*	0.050	4.18E-124	0.036	3.25E-18
cg25949550	*7:145814306*	*CNTNAP2*	-0.014	1.68E-119	-0.010	2.97E-09
cg09935388	*1:92947588*	*GFI1*	-0.099	6.61E-111	-0.021	2.97E-05
cg14179389	*1:92947961*	*GFI1*	-0.068	2.79E-101	-0.039	5.46E-16
cg22132788	*7:45002486*	*MYO1G*	0.023	2.85E-100	0.045	1.86E-20
cg12876356	*1:92946700*	*GFI1*	-0.094	8.35E-99	-0.010	7.28E-03
cg18146737	*1:92946700*	*GFI1*	-0.094	6.14E-91	0.000	7.42E-01
cg19089201	*7:45002287*	*MYO1G*	0.019	3.56E-78	0.043	1.33E-20

The location of these CpGs was predominantly in CpG islands shores, enhancers and DNase hypersensitivity sites while it was found a depletion in CpG islands and promoters. Other studies have been able to replicate these findings in same loci despite of having smaller sample sizes.[145,146]

The 6073 CpGs associated with sustained smoking in Joubert et al were analyzed in older children (7 years) finding a 73% of consistence in effect's direction and also nominal significance for all of them, although 61% showed a weaker effect. This in combination with results from other studies,[140,147] demonstrated a persistence of the dysregulation of some CpGs mapping to *CNTNAP2*, *MYO1G* and *CYP1A1* in childhood (Figure 15). Furthermore, 17 CpGs have been found to be persistently affected by MSDP until midlife (around 43 years), mapping to similar genes (*CYP1A1*, *MYO1G*, *AHRR* and *GFI1)*.[148]

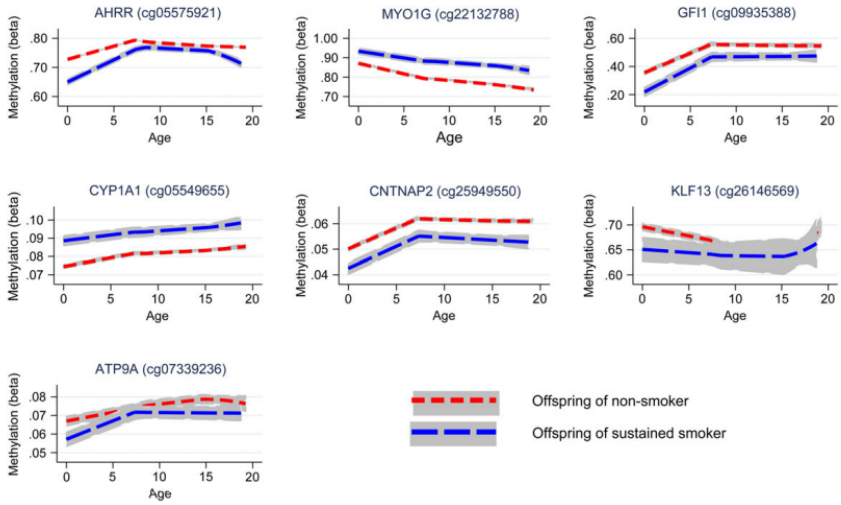

Figure 15. Longitudinal trajectories of DNA methylation of the offspring of smoker and non-smoker mothers showing a persistent effects for some of them.[140]

Comparing to current smoking, DNA methylation differences in CpGs located in *AHRR*, *MYO1G* and *GFI1* associated with MSDP show less pronounced effects.[132] However, some CpGs were only detected in cord blood in relation to MSDP [cg23067299 (*AHRR*) or cg05549655 (*CYP1A1*)], whereas other CpGs are not reported in cord blood despite being identified in current smokers [cg03636183 (*F2RL3*) and cg19859270 (*GPR15*)]. These discrepancies could be explained, among others, by the differences of exposure pathways between pregnancy and current smoking. Current smokers inhale tobacco smoke in a direct way which lead to irritation and inflammation on airways whereas fetus receive it through maternal blood.[149]

DNA methylation is time and tissue specific and in order to elucidate tissue-specific responses to MSDP, a study have analyzed buccal epithelium and placenta.[150]

In 2016, our group published the biggest study to date of the effects of MSDP on placental methylome using the HumanMethylation450 (450K) BeadChip. The study was done in 179 samples from the Child and Environment (INMA) birth cohort. Fifty CpGs were differentially methylated at 5% FDR located in 46 different loci. The top CpG, cg27402634, located between *LINC00086* and *LEKR1* exhibited -16.7% of methylation and seemed to mediate 36.5% of the effect of MSDP on birth weight. Significant CpGs showed an enrichment for pathways related to signaling by growth factors, inflammation, oxidative stress, phagocytosis, peroxisomal metabolism and myometrial relaxation and contraction.[151] Other smaller studies in placenta are summarized in Table 3.[152–159] Some of them also found correlations between placental methylation levels and fetal growth and birth weight.[151,154]

Table 3. Summary of studies assessing placental DNA methylation differences in relation to MSDP.

First author (year)	Sample size	Method	Sig. genes
Suter (2010)	34	Bisulfite pyrosequencing	*CYP1A1*
Suter (2011)	36	27K; bisulfite pyrosequencing	*PURA, GTF2H2, GCA, GPR135, HKR1*
Wilhelm-Benartzi (2012)	380	Bisulfite pyrosequencing; 27K	*LINE-1; AluYb8*
Appleton (2013)	444	Bisulfite pyrosequencing	*HSD11B2*

Maccani (2013)	206	27K; bisulfite pyrosequencing	*RUNX3*
Paquette (2013)	444	Bisulfite pyrosequencing	*HTR2A*
Chhabra (2014)	80	450K	*GTF2H2C, GTF2H2D*
Stroud (2016)	45	Bisulfite pyrosequencing	*NR3C1*
Morales (2016)	179	450K; Bisulfite pyrosequencing	*LINC00086* and *LEKR1, WBP1L, TRIO* among others

1.3.2 miRNA

Some studies have explored the role of miRNAs in subjects actively or passively exposed to smoking in different tissues or cells: spermatozoa,[160] small airway epithelium,[161] plasma,[162] or whole blood (Table 4).[163]

Table 4. Summary of studies assessing miRNA expression levels in relation to smoking

First author (year)	Smoking	Tissue/Cells	Sample size	Method	Num. Sig. miRNA
Marczylo (2012)	Current	Spermatozoa	13	miRCURY LNA	28
Wang (2015)	Current	Small airway epithelium	19	miRNA 2.0 arrays	34
Opstad (2017)	Current	Plasma	1001	Real-time PCR: 1 miRNA	1
Willinger (2017)	Current/ former	Whole blood	5023	TaqMan microRNA assays	6
Maccani (2010)	Maternal smoking	Placenta	25	Real-time PCR: 4 miRNAs	3
Herberth (2013)	Maternal smoking	Cord blood/maternal blood	629	Real-time PCR: 2 miRNAs	1

A study of the expression levels of 754 miRNA in whole blood of 5023 participants, including smokers, former smokers and never smokers, found 5 miRNAs (miR-1180, miR-181a-2-3p, miR-25-5p, miR-1285-3p, miR-342-5p) out of the 289 detectable miRNAs to be associated with current smoking and one with former smoking (miR-423-5p). The expression of these 6 miRNAs correlated with the expression of 116 genes involved in inflammatory pathways in the same blood samples. Moreover, these miRNAs were also related to CRP and pulmonary function.[163,164]

MSDP also seems to modify miRNA expression in placenta and cord blood. Expression of miR-12, miR-21 and miR-146a were significantly downregulated in placenta,[165] whereas expression of miR223 was upregulated in cord blood and maternal blood.[166] High expression of miR-223 was associated with lower T-reg cell number and with a higher risk of atopic dermatitis later in childhood.[165,166]

1.3.3 Transcriptome

The transcriptome of small airway epithelium has been widely studied, as it is the candidate tissue of smoke exposure. Genes differently expressed are involved in immune system regulation, apoptosis and general cellular processes.[167–169] Some studies have demonstrated that the same set of genes is dysregulated in smokers in several proximal tissues including buccal and nasal epithelium,

suggesting that these less invasive tissues might be used as proxies of the lung.[170–172]

Two studies have investigated gene expression changes associated with current smoking in whole blood. The first one (N=209) found 4214 genes differentially expressed at nominal significance involved in apoptosis, cell death, regulation of immune system and response to organic substances. The top genes, were *LRRN3* (leucine-rich repeat protein) and *CLDND1* (claudin-domain containing gene), both upregulated in smokers, which showed some recovery over time after cessation.[173] The second study included 10233 individuals, current, former and never smokers from six cohorts.[174] 1270 genes were differentially expressed at 10% FDR, 56% of which were up-regulated. Top 10 genes are shown at Table 5. Some of these genes were located near CpGs that were found to be hypomethylated in smokers such as *LRRN3* or *GPR15*. Moreover, the expression 19 genes showed persistent effects, 12 of which remained altered 30 years after smoking cessation including *LRRN3, GPR15* and *CLDND1*. Pathway analysis revealed that these genes were mainly involved in platelet activation, lymphocyte activation, inflammatory response, and protein biosynthesis.

Table 5. Top 10 genes differentially expressed in whole blood between smokers and non-smokers.[174]

Gene	Effect	SE	P. Value	FDR
LRRN3	0.64	0.02	1.17E-281	2.94E-277
SASH1	0.18	0.01	1.09E-98	1.36E-94
CLDND1	0.18	0.01	1.00E-66	8.40E-63
PID1	0.25	0.02	1.67E-53	1.05E-49
CLEC10A	0.09	0.01	6.00E-51	3.01E-47
MAL	0.17	0.01	9.34E-35	3.91E-31
PYHIN1	-0.10	0.01	2.42E-34	8.68E-31
CX3CR1	-0.16	0.01	9.49E-33	2.98E-29
GPR15	0.09	0.01	3.15E-30	7.91E-27
TTC38	-0.16	0.01	2.86E-30	7.91E-27

The placental transcriptome is also associated with MSDP.[175] Among the 174 dysregulated genes, *CYP1A1* expression was increased and steroid hormone-metabolizing enzymes were decreased. Nothing is known about the effects of MSDP on offspring gene expression in cord blood or peripheral blood in childhood.

1.3.4 Proteome

Most of the studies investigating differences in protein levels in relation to smoking have analyzed candidate proteins in several cell types or tissues.

C-reactive protein (CRP), a common marker of general inflammatory, shows a dose relationship with current smoking in different tissues.[176–180] Some studies also suggest that CRP remains with high levels after smoking cessation.[181]

An attempt to find an expression profile of urinary proteins in relation to smoking status revealed that most of the proteins affected responded to inflammatory stimulus.[182]

In bronchoalveolar lavage cells, smoking alters the levels of more than 500 proteins, which are involved in 15 molecular pathways including phagosomal and leukocyte trans endothelial migration which are correlated with airway obstruction.[183]

MSDP has been associated with both maternal and fetal protein levels. Maternal serum concentrations of pregnancy-associated plasma protein A (PAPP-A) and insulin-like growth factors (IGF) were found to be lower in active smoker mothers. These proteins are involved in fetal growth regulation during pregnancy and can influence birth weight.[184] Moreover, an increasing in levels of pro-inflammatory proteins has been found in cord blood from smoker mothers.[185] In placenta, 70 proteins were found to be affected by maternal smoking.[186] They were involved in cell morphology, cellular assembly and organization, DNA replication, recombination, repair, energy production and nucleic acid metabolism. Finally, expression of certain proteins in fetal liver was also dysregulated in fetuses exposed to tobacco smoking.[187]

1.3.5 Metabolome

Metabolomics can be used to measure both, exogenous products derived from environmental exposures or their metabolism in the body, as well as endogenous products from physiological processes or from homeostatic responses to toxic exposures. In the case of tobacco smoking, metabolomics can detect metabolites which are products of cigarette compound metabolism as well as changes in endogenous metabolite levels in response to these compounds.

1.3.5.1 Metabolome: exogenous products

Some metabolites that come from cigarette smoking composition can be used as biomarkers of this exposure such as nicotine, cotinine, carbon monoxide, carboxyhemoglobin or thiocyanate, and they can be measured in different specimens as is summarized in Figure 16. Thiocyanate, which is a metabolite of the hydrogen cyanide, it is difficult to quantify and some other metabolites are expensive or invasive.[63] Carbon monoxide and carboxyhemoglobin are cheap to measure and noninvasive but its half-life is very short.

Figure 16. Methods for determining cigarette smoke exposure.[63]

		Properties						
Biomarker	Specimens	Long Half-Life	Inexpensive	Noninvasive	Bedside Test Available	Sensitive	Covers Large Retrospective Period	Not Affected by Nicotine Replacement Therapy
Nicotine	Serum[25]							
	Saliva[25]			✓				
	Urine[25]							
	Hair[24,31]			✓			✓	
	Toe nails[32]						✓	
Cotinine	Serum[25,30]	✓				✓		
	Saliva[33,34]	✓		✓				
	Urine[35,36]	✓			✓			
	Hair[24]	✓		✓			✓	
Carbon monoxide	Exhaled air[35,37-39]		✓	✓	✓			✓
Carboxyhemoglobin	Serum[37]							✓
Thiocyanate	Serum[34]	✓						✓

Cotinine, which the main metabolite of nicotine, is the most popular biomarker of exposure to tobacco smoking. In humans, about 70-80% (Figure 17) of nicotine is metabolized to cotinine in the liver, mainly by cytochrome P450 2A6 (CYP2A6).[188] It is the most widely used biomarker for two reasons: i) it can be measured in different biological samples (urine, blood, hair or saliva) using different techniques and ii) its half-life (approximately 20h) is higher than nicotine (2-3 hours) and its concentration is directly proportional to the exposure to nicotine.[189,190] It is mainly measured in urine due to its easily measurement and the availability of a test. In other specimens, it is more expensive, invasive in some cases or the process needs to be standardized.

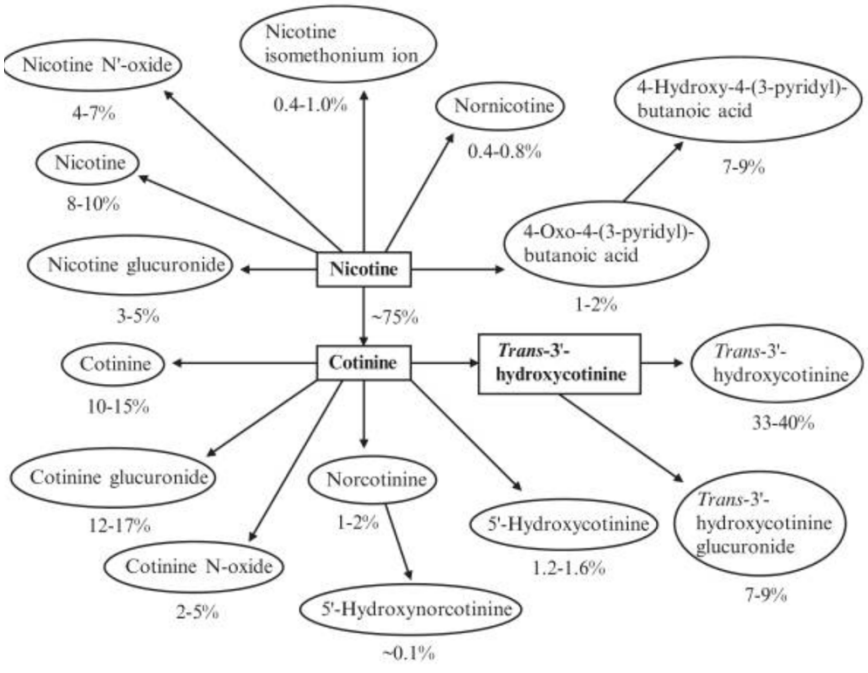

Figure 17. Nicotine is metabolized in the liver to cotinine (nearly 80%) by a group of cytochrome enzymes.[191]

Although cotinine is the best biomarker of tobacco smoking, it presents some disadvantages: i) its half-life is relatively short (about 16h hours in adults[191] and 10h in pregnant women)[190] informing only about recent exposure to tobacco smoking, ii) it depends on the detoxification efficiency of CYP2A6 and UGT1A enzymes which show gender or ethnic differences and iii) it does not allow a clear discrimination of unexposed and SHS.[190,192,193]

1.3.5.2 Metabolome: endogenous products

Several studies have tried to identify biomarkers, exogenous or endogenous metabolites, of tobacco exposure with better properties than cotinine in different biological matrices (saliva,[194] urine,[195–197], serum[198] and plasma).[199] However, further investigation is needed in this field.

On the other hand, Xu, *et al.* investigated the metabolic consequences of current smoking in 1036 adults.[200] Serum metabolome was measured with the Biocrates platform (AbsoluteIDQ P180 and P150 kits). 18 metabolites were identified in men and 6 in women significantly altered between smokers and non-smokers (Figure 17). Only few metabolites showed persistent effects after a maximum of 7 years after smoking cessation, being the major part reversible. In smokers, they found higher unsaturated diacyl-phosphatydylcholines (aa-PCs), but lower acyl-alkyl-PCs (ae-PCs) and saturated aa-PCs. Although it is not clear the role of these metabolic alterations,

unsaturated fatty acids can influence the risk of different diseases including cardiovascular diseases.[201,202] Arginine and glutamate were also increased in smokers and they have been also found to be associated with cardiovascular diseases.[203]

Metabolites	Current smokers vs Non-smokers	
	Odds ratio (95% CI)	P. Value
Men	**125 vs 200**	
Arginine	1.7 (1.3, 2.2)	2.6E-05
Aspartate	1.6 (1.2, 2.0)	2.5E-04
Glutamate	1.6 (1.2, 2.0)	6.2E-04
Ornithine	1.4 (1.2, 1.9)	2.2E-03
Serine	1.4 (1.1, 1.8)	3.5E-03
Kynurenine	0.6 (0.5, 0.9)	3.2E-03
PC aa C32:3	0.7 (0.5, 0.9)	6.4E-03
PC aa C34:1	1.7 (1.3, 2.2)	2.0E-04
PC aa C36:0	0.6 (0.5, 0.8)	3.5E-04
PC aa C36:1	1.6 (1.2, 2.0)	9.4E-04
PC aa C38:0	0.7 (0.5, 0.9)	2.1E-03
PC aa C38:3	1.5 (1.1, 1.9)	3.4E-03
PC aa C40:4	1.5 (1.2, 2.0)	3.4E-03
PC ae C34:3	0.5 (0.4, 0.7)	3.3E-06
PC ae C38:0	0.7 (0.5, 0.9)	2.1E-03
PC ae C38:6	0.7 (0.5, 0.9)	4.8E-03
PC ae C40:6	0.6 (0.5, 0.8)	8.8E-04
lysoPC a C18:2	0.7 (0.5, 0.9)	3.3E-03
Women	**70 vs 205**	
Carnitine	1.8 (1.4, 2.4)	4.3E-05
Glutamate	1.7 (1.3, 2.2)	1.2E-04
PC aa C32:1	1.5 (1.1, 1.9)	2.1E-03
PC aa C36:1	1.6 (1.2, 2.0)	1.1E-03
PC ae C34:3	0.6 (0.4, 0.8)	7.7E-04
SM (OH) C22:2	0.6 (0.5, 0.8)	2.1E-03

Figure 18. Serum metabolites significantly dysregulated at FDR 5% in current smokers (CS) in comparison to non-smokers. Results from Xu, *et. al.* study.[200]

2 HYPOTHESIS AND OBJECTIVES

Tobacco smoking is one of the leading causes of preventable death and disability. The prenatal period and the first years of life are critical periods for child development, and environmental impacts during these years can lead to short- and long-term adverse consequences for health. Despite knowing about the negative health outcomes of exposure to tobacco smoking, the molecular mechanisms are only partially understood.

In this thesis, it has been worked on the hypothesis that exposure to tobacco smoking can alter epigenetic patterns and other molecular phenotypes in different tissues and that some of them can remain dysregulated during long periods of time.

The main objective of this thesis is to investigate the association of both maternal tobacco smoking during pregnancy (MSDP) and prenatal and postnatal exposure to secondhand smoking (SHS) with different molecular phenotypes in children. The particular aims are:

1. To study the association between MSDP and placental DNA methylation, its link with reproductive outcomes, and the overlap with smoking signatures reported in the cord blood methylome.

2. To study the association of exposure to tobacco smoking in two different periods, during pregnancy and postnatally, with different layers of molecular phenotypes measured in childhood, including blood DNA methylation and

transcription, plasma proteins and serum and urinary metabolites.

With these two objectives we plan to address some fundamental questions in molecular environmental epidemiology: tissue specificity, susceptibility time windows, and dose-response trends.

3 MATERNAL TOBACCO SMOKING DURING PREGNANCY AND PLACENTAL DNA METHYLATION, AND ITS LINK WITH REPRODUCTIVE OUTCOMES

3.1 Materials and methods

3.1.1 Study population

Cohorts that are members of the PACE consortium were identified for participation in the current study if they had existing DNA methylation data quantified from placental tissue via the Illumina Infinium HumanMethylation450 BeadChip and if they had obtained information of self-reported maternal smoking during pregnancy (MSDP). The seven cohorts that contributed to the meta-analysis of any-MSDP included Asking Questions about Alcohol in pregnancy (AQUA),[204] Study on the pre- and early postnatal determinants of child health and development (EDEN),[205] Genetics of Glucose regulation in Gestation and Growth (Gen3G),[206] Genetics, Early Life Environmental Exposures and Infant Development in Andalusia (GENEIDA),[207] Environment and Childhood Project (INMA),[208] New Hampshire Birth Cohort Study (NHBCS)[209] and Rhode Island Child Health Study (RICHS)[210] (Figure 19).

EDEN, GENEIDA and INMA also contributed to the sustained-MSDP analyses. RICHS contributed RNA-seq data for analyses with gene expression. All cohorts acquired ethics approval and informed consent from participants prior to data collection through local ethics committees. All participants in the study were European ancestry, except 1.85% of EDEN mothers, which were not native.

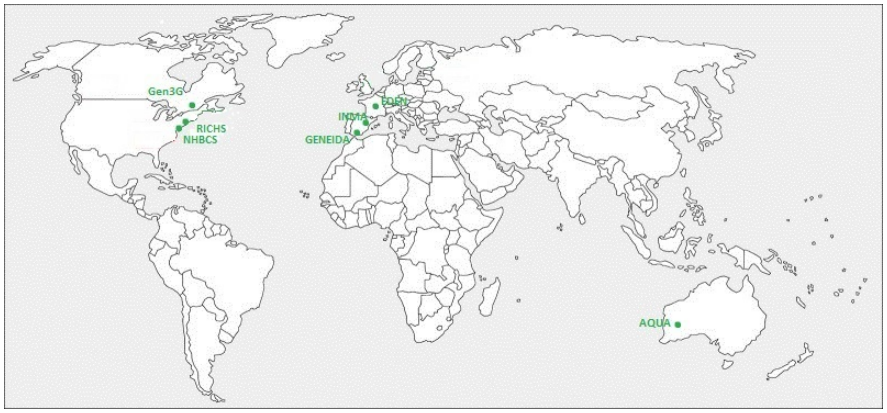

Figure 19. Participating cohorts

3.1.2 Tobacco smoking definitions

Any-MSDP (A-MSDP was defined as "yes" if mothers reported smoking cigarettes at any time during pregnancy. Sustained-MSDP (S-MSDP) was defined as "yes" when mothers reported smoking cigarettes at least in the 1st and 3rd trimester of pregnancy. For both exposure variables, the comparison group was defined as mothers that reported no smoking during pregnancy.

3.1.3 Placental genome-wide DNA methylation data

Placental DNA METHYLATION from the fetal site was assessed with the Infinium Human-Methylation 450K array (Illumina, San Diego, CA USA). Quality control of DNA METHYLATION was standardized across all cohorts. Low quality samples were filtered out and probes with detection P. Values > 0.01 were excluded. Beta-values were normalized via functional normalization[211] and beta-

mixture quantile normalization (BMIQ)[212] was applied to correct for the probe type bias. Cohorts applied ComBat to remove batch effects when applicable. Probes that hybridize to the X/Y chromosomes, cross-hybridizing probes and probes with SNPs at the CpG site, extension site, or within 10 bp of the extension site with an average minor allele frequency > 0.01 were filtered out.[213] Overall, 418639 probes and 415396 probes were available for A-MSDP and S-MSDP models. Finally, methylation extreme outliers (<25th percentile – 3*IQR or 75th percentile +3*IQR across all the samples) were trimmed.

3.1.4 Estimates of putative cellular heterogeneity

Placental cellular heterogeneity was estimated from DNA METHYLATION data using a reference-free cell-mixture decomposition method.[214] The number of surrogate variables ranged from 2 to 5 depending on the cohort. Models for differential methylation were corrected for the number of surrogate variables minus one to reduce multi-collinearity.

3.1.5 Genome-wide differential DNA methylation analyses

Within each cohort, robust linear regression from the "MASS" package)[215] in R were used to account for potential heteroskedasticity while testing the associations between normalized methylation beta values at each CpG with A-MSDP and S-MSDP. Models were

adjusted for maternal age, parity, maternal education and cohort-specific variables first unadjusted for putative cellular heterogeneity (models A1 and S1) then adjusted for cellular heterogeneity (models A2 and S2).

After checking results sent by the cohorts, we conducted inverse variance-weighted fixed-effects meta-analyses using METAL[216] to estimate pooled associations. The meta-analysis was performed independently by two groups to ensure consistent results. CpGs not present in at least 2 cohorts and at least in >10% of smokers were filtered out. Forest plots were creating using the "meta" R package, and Volcano plots with "calibrate" R package. To examine whether increased duration of exposure (sustained smoking versus any smoking) yielded increased magnitudes of association, we calculated the percent change in the coefficients between the two models ($|\beta_{A2}| - |\beta_{S2}|)/|\beta_{A2}| * 100$. We provide the results for all models yielding associations within a 5% FDR, but only perform downstream analyses on CpGs that yielded Bonferroni significant associations with MSDP in models adjusted for cellular heterogeneity, models A2 (N=878) and S2 (N=894).

3.1.6 Expression quantitative trait methylation (eQTM) loci

We performed expression quantitative trait methylation (eQTM)[103] in RICHS cohort. Transcription was measured via RNA-seq on 194 placentas. The details of sample collection, assay, and QC for the RNA-seq data are presented in detail elsewhere.[217] In this dataset, we

identified 6523 unique transcripts annotated to an Ensemble ID (GrCh37/hg19) and with a transcriptional start site (TSS) within 250 kb upstream or downstream of 1184 out of the 1224 candidate CpGs. The association between DNA METHYLATION and expression levels was assessed via 10295 linear regression models using the MEAL[218] package in R. Because this analysis contained placental samples from an ethically and racially diverse population (23% non-white), we adjusted the models with the first EWAS principal component (PC) calculated with EPISTRUCTURE.[219] We provide the results for all models yielded nominally significant associations (raw P. Values <0.05), statistically significant eQTMs were determined at 5% FDR.

3.1.7 CpG site annotation

We annotated CpGs with notations from the Illumina HumanMethylation 450k annotation file,[220] and with several regulatory features: placental 15-chromatin states[221] released from the ROADMAP Epigenomics Mapping Consortium (ChromHMM v1.10),[84] placental partially methylated domains (PMDs)[112] and placental germline differentially methylated regions (gDMRs).[222]

3.1.8 Enrichment analyses

Functional enrichment analyses were performed via ConsensusPathDB[223] using KEGG, Reactome, Wikipathways,

Biocarta as reference gene-sets. ConsensusPathDB performs a hypergeometric test and corrects multiple-testing with the FDR method. Enrichment for transcription factors and for phenotypes were assessed with EnrichR using ENCODE and ChEA consensus TFs from ChIP-X database, and dbGaP database, respectively. EnrichR results were ranked using the combined score (P. Value computed using Fisher exact test combined with the z-score of the deviation from the expected rank).[224] Enrichment for regulatory features was assessed with the hypergeometric test, and P. Values were corrected with BN considering 15 (placental chromatin 15-states) or 6 (relation to CpG island) elements.

3.1.9 Overlap of MSDP-sensitive CpG sites and birth weight loci

The overlap between MSDP-sensitive CpGs in placenta and maternal/fetal BW, BL, HC and GA loci was done using the largest genome-wide association studies (GWAS) to date.[225–229] First, a list of independent single nucleotide polymorphisms (SNPs) related to reproductive outcomes was obtained considering linkage disequilibrium patterns and information from the manuscripts. Then, we identified which CpGs were located in 1Mb surrounding each of the 74 independent SNPs (0.5Mb in each side of the CpG).

3.1.10 Association between DNA methylation and reproductive outcomes

Within each cohort, robust linear regression models were utilized (MASS R package)[230] to test the association between normalized methylation beta values at each CpG as the independent variable and gestational duration (inverse normal transformation of sex residuals, in days), pre-term birth (defined as <37 weeks of gestation), birth weight z-scores, birth length z-scores, and head circumference z-scores as the dependent variables. Birth size z-scores were calculated using international references from the INTERGROWTH-21st Project[231] and standardized by both gestational age and fetal sex. Models were adjusted for maternal age, parity, maternal education and cohort-specific variables, and for putative cellular heterogeneity. Inverse variance-weighted fixed-effects meta-analyses[216] were again used to estimate pooled associations. Multiple testing was controlled with the Bonferroni adjustment.

3.1.11 Comparison of MSDP-sensitive CpGs in placenta with cord blood

We examined the consistency between MSDP-sensitive CpGs in placenta and in cord blood[138] using Pearson correlation coefficients. We compared the coefficients from the models for sustained MSDP in cord blood to results for both any (A2) and sustained (S2) MSDP in placenta.

3.2 Results

3.2.1 Study population

A total of seven cohorts (N=1700) contributed to the epigenome-wide association study (EWAS) of any-MSDP, including AQUA,[232] EDEN,[205] Gen3G,[206] GENEIDA,[207] INMA,[208] NHBCS[209] and RICHS.[210] For this primary meta-analysis, 20.24% of mothers reported ever smoking during pregnancy, defined as any self-reported cigarette smoking during any trimester of pregnancy. MSDP tended to be less prevalent in the cohorts from Canada and the USA compared to those from Australia and Europe (Table 6).

Table 6. Cohorts and percentage of any smoker mothers during pregnancy.

Cohort	Country	A-MSDP		
		Total N	N (%) Non-smokers	N (%) Smokers
AQUA	Australia	99	75 (75.76)	24 (24.24)
EDEN	France	647	466 (68.93)	201 (31.07)
Gen3G	Canada	151	138 (91.39)	13 (8.61)
GENEIDA	Spain	87	67 (77.01)	20 (22.99)
INMA	Spain	166	119 (71.69)	47 (28.31)
NHBCS	USA	310	290 (93.55)	20 (6.45)
RICHS	USA	240	221 (92.08)	19 (7.92)
TOTAL		1700	1356 (79.76)	344 (20.24)

Three cohorts (N=795), EDEN, GENEIDA and INMA, contributed to the EWAS of sustained-MSDP, among which 20.25% of mothers reported smoking throughout pregnancy (Table 7). Distributions of covariates by cohort are provided in Supplemental Excel Table S1).

Table 7. Cohorts and percentage of sustained smoker mothers during pregnancy

Cohort	Country	S-MSDP		
		Total N	N (%) Non-smokers	N (%) Smokers
EDEN	France	570	446 (68.93)	124 (21.75)
GENEIDA	Spain	82	67 (77.01)	15 (18.29)
INMA	Spain	143	119 (71.69)	24 (16.78)
TOTAL		795	632 (79.50)	163 (20.50)

3.2.2 Genome-wide methylation meta-analyses

We produce four statistical models for each CpG site, first unadjusted for putative cellular heterogeneity (models A1 and S1, for any- and sustained-MSDP) and then adjusted for putative cellular heterogeneity (models A2 and S2, for any- and sustained MSDP). All models were adjusted for maternal age, parity and maternal education (except for Gen3G, which not have information about maternal education). Lambda inflation factors from the meta-analyses ranged from 2.00 to 2.84 (Supplemental Excel Table S2, Supplementary Figure S1), did not substantially differ after cellular heterogeneity correction, and were lower among the sustained-MSDP models. Cohorts-specific lambdas tended to correlate with the number of

smokers in each cohort, with EDEN, the largest cohort, showing the largest inflation (λ=1.93). The leave-one-out meta-analysis gave consistent findings when any individual cohort was excluded.

We identified numerous MSDP-associated epigenetic CpGs in the placenta at a 5% false discovery rate (FDR), which are reported in the Supplemental Excel Tables S3-S6. A-MSDP associated with DNA methylation at 532 CpG sites after Bonferroni (BN) adjustment (Supplemental Excel Table S3), while S-MSDP associated with DNA methylation at 568 CpG sites (Supplemental Excel Table S4). After adjusting for putative cellular heterogeneity, A-MSDP associated 878 CpGs (Supplemental Excel Table S5), while S-MSDP associated 894 CpGs (Supplemental Excel Table S6) after BN adjustment. The summary of these results is shown in Table 8.

Table 8. Summary of results from genome-wide methylation meta-analysis

Model	Exposure	Adj. cellular heterogeneity	Total N CpGs	5% FDR CpGs	BN CpGs
A1	A-MSDP	No	418639	25968	532
A2	A-MSDP	Yes	418639	31825	878
S1	S-MSDP	No	415416	11234	568
S2	S-MSDP	Yes	415396	19218	894

Heterogeneity in the associations across cohorts was lower among EWAS that were adjusted for putative cellular heterogeneity. For example, among the BN-significant sites 10.7% of A1 hits and 5.5% of A2 hits yielded heterogeneity P. Values <0.01, while only 3.9% of S1 and 3.5% of S2 hits produced heterogeneity P. Values < 0.01.

Thus, the BN-significant results adjusted for cellular heterogeneity (A2 and S2) were carried forward for secondary analyses.

Overall, 67% of our BN-significant hits exhibited lower DNA methylation associated with MSDP (Figure 20).

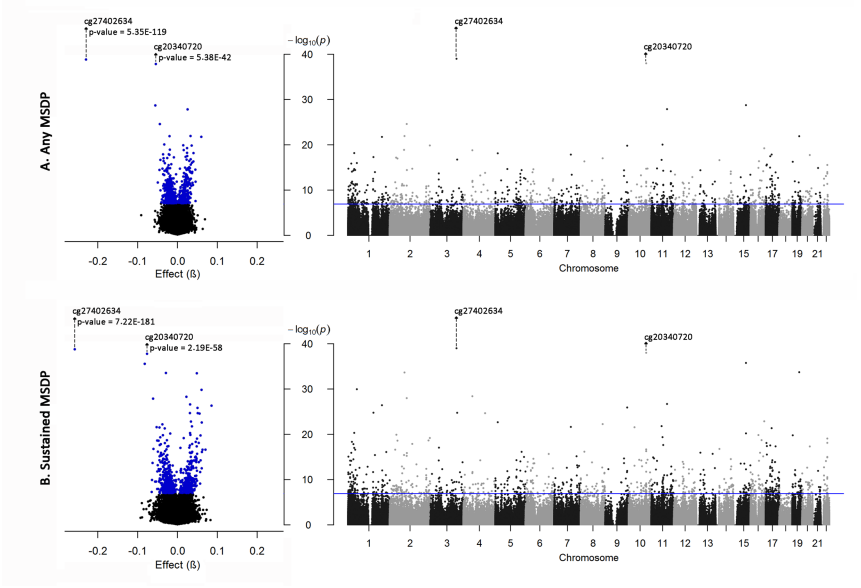

Figure 20. Volcano and Manhattan plots of the association between any MSDP (N=1700, exposed = 344) (A) and sustained MSDP (N=795, exposed = 163) (B) with placental DNA methylation adjusted for maternal age, parity, maternal education and putative cellular heterogeneity. In the volcano plots, the x-axis shows the difference in DNA methylation (effect) with a possible range between 0 and 1, while the x-axis in the Manhattan plot represents genomic location; both plots share the same y-axis with $-\log_{10}(P)$. Bonferroni thresholds for statistical significance are shown as blue dots and a blue horizontal line, for volcano and Manhattan plots respectively. The y-axes were truncated to a minimum p-value of 1E-40 (or maximum $-\log_{10}(P)$ of 40), to a allow for better visualization of the majority of our results. The CpGs that were impacted by y-axis truncation are indicated with arrows and annotated with their actual p-values.

On average, the absolute mean methylation differences for sustained-MSDP (Δ = 2.8%) was greater than for any-MSDP (Δ = 1.9%), and

we found that all 548 CpGs that were BN-significant in both A2 and S2 models, had larger estimated magnitudes of the associations with increased duration of exposure (Supplemental Excel Table S7). However, data from only three cohorts were available to obtain pooled estimates of association for model S2. Thus, we examined whether the increase in magnitudes of associations between models S2 and A2 were reproducible when A2 models were restricted to EDEN, GENEIDA and INMA. We found that all but 4 loci (544 out of 548) again exhibited increased magnitudes of association with increased duration of exposure (Supplemental Excel Table S7, Supplementary Figure S2).

The most notable association from our analysis was observed at cg27402634, which is located between a long non-coding RNA (*LINC00086*) and the *Leucine, glutamate and lysine rich 1 (LEKR1*) gene, showed the largest differential DNA methylation. Mothers that smoked any cigarettes during pregnancy had 22.95% lower DNA methylation (P. Value = 5.35E-119) and mothers that smoked throughout pregnancy had 25.78% lower DNA methylation (P. Value = 7.22E-181) when compared to mothers that did not smoke at all during pregnancy. Cohort-specific differential DNA methylation varied substantially at this locus for any-MSDP, with NHBCS exhibiting the greatest hypomethylation (-41.24%), and AQUA the least (-7.17%), while we did not detect substantial heterogeneity among cohort-specific associations from the sustained-MSDP analysis (Figure 21A). The random effects meta-analyses for cg27402634 yielded similar associations but larger P. Values, due to heterogeneity between cohorts, for any (β = -19.79%, P. Value =

7.47E-06) and sustained (β = -25.25%, P. Value = 6.13E-63) MSDP, respectively. We observed striking consistency in the associations across cohorts for the other BN-significant CpGs that yielded magnitudes of association ($|\beta_{A2}|$) > 0.05 for cg26843110, cg20340720 and cg17823829 (Figure 21B-21D).

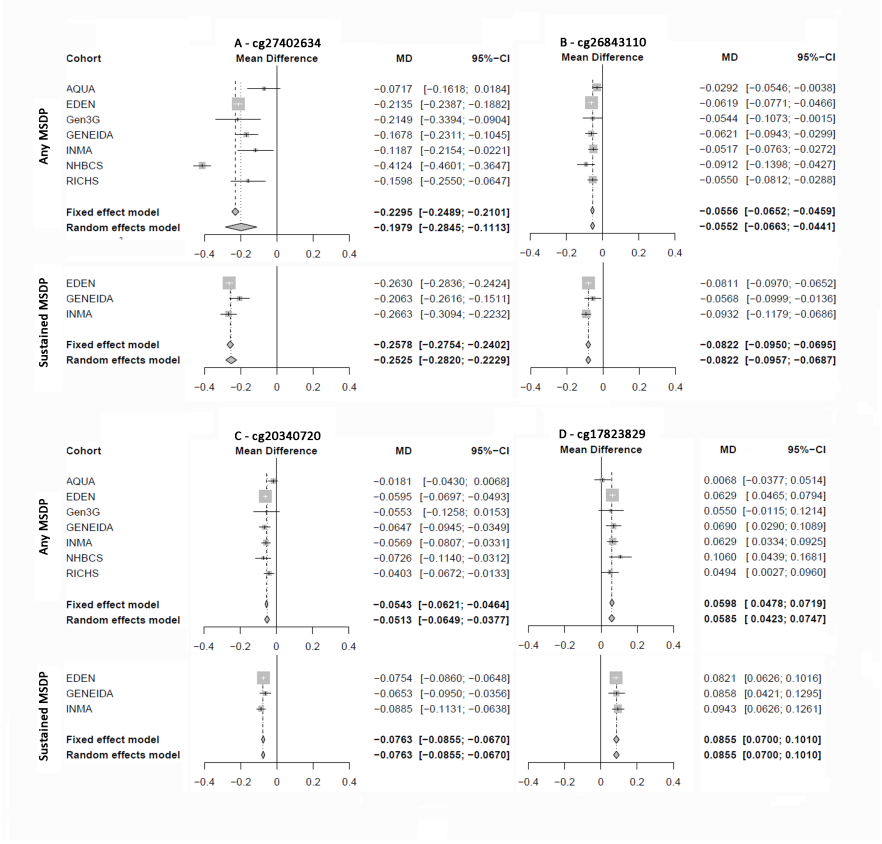

Figure 21. Forest plots of the pooled and cohort specific estimates of association between placental DNA methylation levels at (A) cg27402634, (B) cg26843110, (C) cg20340720, and (D) cg17823829 with any MSDP (A2) and sustained MSDP (S2); models adjusted for maternal age, parity, maternal education, and putative cellular heterogeneity

We identified numerous other noteworthy associations but due to the large number of BN-significant hits from these meta-analyses, we

highlight the relationship among the top 20 CpG sites from the primary meta-analysis of any-MSDP going forward (Table 9) while results for the other CpGs are provided in Supplemental Excel Tables.

3.2.3 Expression Quantitative trait methylation (eQTM) analyses

We tested whether the methylation levels at MSDP-associated CpGs associated with the expression nearby (± 250kb of the CpG) genomic features. For this analysis, we utilized the less conservative false discovery rate (FDR) due to a limited sample size, while aiming to annotate our hits with the genes they most likely have a functional relationship with. We identified 170 associations between methylation and gene expression within 5% FDR (Supplemental Excel Table S8) including 141 unique CpGs and 140 unique RNA transcripts. 72.4% of the eQTMs exhibited inverse associations, while inverse associations were most common and statistical significance was stronger for CpGs that were closer to the TSS (Supplementary Figure S3). Among the top 20 hits from our meta-analysis, eight were identified as eQTMs at a 5% FDR (Table 9). Of particular note, higher placental DNA methylation at cg27402634, the top hit from the meta-analysis, was associated with lower expression (β = -2.32, P. Value = 7.1E-04) of *LEKR1*; this CpG site is ~6 kb upstream

Table 9. Meta-analysis results from models A2 and S2, among top 20 hits from model A2 and results from eQTM models, DNA methylation versus gestational duration models, and DNA methylation versus BW z-score models that yielded at least nominally significant associations. *Nominal statistical significance P. Value < 0.05; ** Within a 5% FDR; *** Bonferroni significance

MarkerName	Any MSDP (Model A2)		Sustained MSDP (Model S2)		% Increase in magnitude (β) from A2 to S2	Top eQTM				DNAM vs. Gestational Duration			DNAM vs. BW Z-score			CpG Annotations			
	Fixed Effect (β)	P. Value	Fixed Effect (β)	P-value		Gene Name	Effect (β)	P. Value	Sig.	Fixed Effect (β)	P. Value	Sig.	Fixed Effect (β)	P. Value	Sig.	Chr	Pos	UCSC_RefGene_Name	UCSC_RefGene_Group
cg00534380	-0.044	2.53E-25	-0.061	1.03E-28	38.46	TBC1D8	-2.292	1.13E-05	**	3.526	9.26E-20	***	1.479	1.62E-04	**	chr2	101766586	*TBC1D8*	Body
cg00666842	0.033	1.17E-22	0.049	2.23E-34	48.17	-	-	-		-	-		-	-		chr2	88366145	*SMYD1*	TSS1500
cg02341503	-0.02	1.81E-18	-0.025	4.83E-18	25.74	ZMYND8	-1.721	7.37E-03	*	1.438	2.29E-02	**	3.185	2.70E-07	***	chr20	45947123	*ZMYND8;LOC100131496*	Body;TSS200
cg03313447	-0.02	1.26E-22	-0.029	1.93E-34	45.73	TGFB1	-10.882	7.27E-13	***	-2.476	7.79E-04	**	2.294	1.58E-03	**	chr19	41829042	*CCDC97*	3'UTR
cg06716730	-0.022	1.4E-18	-0.032	4.47E-22	46.79	DUSP14	-3.073	1.59E-03	*	3.715	7.87E-12	***	-	-		chr17	35851459	*DUSP14*	5'UTR
cg09491670	0.016	1.6E-19	0.023	3.96E-29	43.31	USP46	-3.589	4.91E-04	**	-	-		-	-		chr4	53529646	-	-
cg12291408	-0.036	1.43E-18	-0.052	2.34E-22	45.15	ACTL6B	-4.893	1.12E-02	*	-	-		1.307	6.82E-04	**	chr7	100037572	-	-
cg14214914	0.036	1.57E-20	0.05	1.2E-26	41.29	CRAT	-4.762	6.03E-14	***	-0.96	1.35E-02	**	-	-		chr9	131870304	*CRAT;CRAT*	Body;Body
cg17823829	0.06	1.79E-22	0.086	3.8E-27	42.98	PPFIA4	-2.533	4.05E-04	**	-1.177	9.12E-06	***	-1.021	9.72E-05	**	chr1	202765754	*KDM5B*	Body
cg19246018	0.015	1.39E-20	0.016	6.52E-20	8.28	-	-	-		-1.394	1.74E-02	**	-	-		chr2	240031588	*HDAC4*	Body
cg20340720	-0.054	5.38E-42	-0.076	2.19E-58	40.52	-	-	-		-	-		1.864	2.42E-07	***	chr10	104512523	*C10orf26*	Body
cg23752985	-0.019	1.31E-19	-0.025	3.77E-16	32.98	CAPG	6.326	3.96E-04	**	5.47	1.40E-15	***	-	-		chr2	85803571	*VAMP8*	TSS1500
cg24177452	0.025	3.4E-18	0.031	2.84E-18	27.35	MYO18	-1.373	1.79E-03	*	-	-		-	-		chr17	27494295	*MYO18A*	5'UTR
cg25585967	0.04	7.53E-19	0.061	2.16E-23	51.74	-	-	-		-0.968	8.64E-03	**	-1.348	2.06E-04	**	chr5	14452105	*TRIO*	Body
cg26045080	0.024	6.66E-19	0.032	4.86E-21	31.12	SH3D21	-4.265	8.34E-08	***	-1.407	4.19E-03	**	-	-		chr1	36807363	*STK40*	Body
cg26115089	0.025	1.33E-28	0.032	2E-27	24.8	-	-	-		-1.103	4.43E-02	*	-1.484	6.71E-03	**	chr11	93846406	*HEPHL1*	3'UTR
cg26433445	0.024	5.85E-20	0.034	1.34E-23	40.25	-	-	-		-	-		-2.987	9.06E-07	***	chr16	81764289	-	-
cg26648103	-0.033	8.98E-21	-0.043	4.3E-20	30.51	-	-	-		1.064	1.29E-02	**	1.663	7.56E-05	**	chr11	66791718	*SYT12*	5'UTR
cg26843110	-0.056	1.74E-29	-0.082	1.74E-36	47.84	CSK	0.573	2.18E-02	*	2.295	5.09E-12	***	1.282	1.19E-04	**	chr15	74935742	*EDC3*	Body
cg27402634	-0.23	5.4E-119	-0.258	7.2E-181	12.33	LEKR1	-2.317	7.14E-04	**	-	-		0.922	6.71E-07	***	chr3	156536860	-	-

of the TSSs for both *LEKR1* isoforms. Additionally, one other CpGs upstream of *LEKR1* was BN-significant in our meta-analysis, cg01201279, which was also inversely associated with *LEKR1* expression (β = -2.22, P. Value = 1.96E-03), though not within the 5% FDR threshold.

Also noteworthy, some genes had multiple candidate CpGs that were associated with their expression at a 5% FDR, such as *adenylate cyclase 7* (*ADCY7*) with 4 CpGs, including the top eQTM hit (cg05362860, β = -4.98, P. Value = 1.05E-15), and *Synaptogyrin 1* (*SYBGR1*) and *Sulfotransferase family 2B member 1* (*SULT2B1*), each of which associated with five different CpGs at a 5% FDR.

3.2.4 Functional and regulatory enrichment analyses

We performed functional and regulatory enrichment analyses to gain insights into which biological processes may be influenced by our overall set of MSDP BN-associated CpGs. We first performed gene-set enrichment analyses to identify biological pathways associated with our gene-sets and identified 143 pathways that were significantly (Q. Value < 0.05) enriched in A2 (Supplemental Excel Table S9) and 25 pathways in S2 (Supplemental Excel Table S10). In both gene-sets, "signaling by NGF" was the most significant pathway (Q. Value = 2E-04 and 9.6E-03 for models A2 and S2, respectively). Other significantly enriched pathways related to growth factors (VEGF, EGF, PDGF, IGF1R), hormones (aldosterone, insulin, TSH, GnRH), interleukins (IL-2, IL-4, IL-7),

myometrial contraction, vascular smooth muscle contraction, thrombin and platelet activation, signaling and aggregation. We also tested whether placental MSDP-associated CpGs were enriched for genes regulated by specific transcription factors (TFs). Most notably, our hits were enriched by GATA1 and GATA2, followed by RUNX1, TP63, SMAD4, AR, TP53 or PPARG (Supplemental Excel Tables S11-S12).

We then examined whether the MSDP-associated CpG sites were enriched for partically methylated domains (PMD)[112], CpG island locations, allele specific germline differentially methylated regions (gDMR)[222] or regulatory features from the placenta specific 15-chromatin state annotation from ROADMAP. We found that MSDP-associated CpGs were depleted in PMDs (Supplementary Figure S4), highly enriched in enhancers and depleted in transcription start sites and inactive states (Supplementary Figure S5). We also found MSDP-associated CpGs were depleted in CpG islands and shores, while enriched in CpG islands shelfs and open sea positions (Supplementary Figure S6). We did identify 3 CpGs (cg05211790 and 16360861 at *RAI14*, and cg05575921 at *AHRR* gene) that were within two candidate maternal gDMRs, but our overall set of MSDP-associated CpGs were neither depleted or enriched for gDMRs (Supplementary Figure S7).

3.2.5 Phenotype enrichment analyses

To better understand the pathological processes that may be associated with disrupted regulation of the genes from our MSDP-associated loci, we performed phenotype enrichment analyses. We found that our MSDP-associated CpGs were enriched for numerous phenotypes, including several terms of body shape (body mass index (BMI), waist-hip ratio, obesity), blood pressure, cardiovascular diseases, type 2 diabetes, asthma and respiratory function (Supplemental Excel Tables S13-S14). BMI was the most highly enriched phenotype in both A2 and S2 lists.

3.2.6 Proximity to genetic variants for reproductive outcomes

We also aimed to understand whether the genetic influences on reproductive outcomes were located in similar genomic regions to our MSDP-associated CpGs. Thus we investigated whether 74 maternal/fetal single nucleotide polymorphisms (SNPs) that have been associated with birth weight (BW),[225,226] birth length (BL)[227], head circumference (HC)[228] and gestational age (GA)[229] were in similar genomic regions (within 1Mb) to MSDP-associated CpGs, and have been added as annotations to the meta-analysis results files (Supplemental Excel Tables S3-S6). Of the 71 independent SNPs in autosomal chromosomes, 22 and 21 were within 1 Mb of any- or sustained-MSDP-associated CpGs, respectively (Supplemental Excel Tables S15-S16). Two out of the top four CpGs were close to

BW related SNPs. Cg27402634 was within ~250 kb and ~707 kb of the fetal-effects rs13322435[225] and maternal-effects rs9872556,[226] respectively; and cg20340720 was within ~401 kb of the fetal-effects rs74233809.[225] It is worth nothing that all but one of the 71 autosomal SNPs were within 1 Mb of MSDP-associated CpGs that were within a 5% FDR.

3.2.7 Association of DNA methylation and reproductive outcomes

In addition to examining whether our hits were in close proximity to genetic variants that have previously been associated with reproductive outcomes, we aimed to understand how our MSDP-associated loci themselves were associated fetal growth and development. Thus, we examined the relationships between DNA methylation with gestational duration, preterm birth, and birth size (BW, BL and HC z-scores) at our set of 1224 CpG sites that associated with MSDP. We found that 341 out of the 1224 CpGs (27.9%) were related to at least one reproductive outcome after Bonferroni-adjustment. The majority of these (298 CpGs) were associated with gestational duration (Supplemental Excel Table S17). Preterm delivery, for which only two cohorts could contribute data (EDEN and NHBCS), produced similar associations found for GA, though fewer CpGs were statistically significant (Supplemental Excel Table S18). The top CpG associated with gestational age was cg00534380 ($\beta = 3.53$, P. Value = 9.26E-20), which is located within

the gene body of *TBC1D8* (*TBC1 Domain Family Member 8*). Preterm risk increased in 16% per each 1% methylation reduction at this CpG site (P. Value = 7.17E-09).

We also found that numerous loci were associated with birth size z-scores after Bonferroni-adjustment, with the majority of these being associated with birth weight (43 CpGs) (Supplemental Excel Table S19), followed by BL (20 CpGs) (Supplemental Excel Table S20) and HC (4 CpGs) (Supplemental Excel Table S21). Some of the CpGs associated with GA were also associated with birth size measurements, even though they were standardized for GA, suggesting independent associations with both gestational duration and fetal growth (Figure 22A), including 6 CpGs (annotated to *SYNJ2, PXN, PTPRE, IGF2BP2* and 4q21.1) shared between GA and BW, and 2 (annotated to *POLR3E* and *LOC441869*) shared between GA and BL. Among the CpGs that yielded at least one BN-significant association with reproductive outcomes, CpGs that tended to have positive associations with reproductive outcomes clustered together and were typically hypomethylated with MSDP, while CpGs that exhibited inverse associations with reproductive outcomes tended to be hypermethylated with MSDP (Figure 22B).

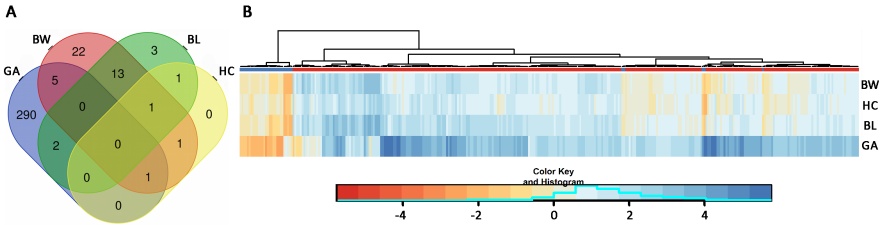

Figure 22. (A) Venn diagram of the CpGs associated to the reproductive outcomes (GA = Gestational Age, BW = Birth Weight, BL = Birth Length and HC = Head Circumference) at BN-threshold for significance, and (B) Clustering of coefficients for GA, BW, BL, and HC among the CpGs that yielded a BN-significant association for at least one reproductive outcome (blue = positive, red = inverse); top axis color bar shows direction of association between DNA methylation and MSDP (blue = positive, red = inverse) from the meta-analysis results. In general, high methylation level in MSDP hypermethylated CpGs was associated with reduced GA and birth size; while high methylation level in MSDP hypomethylated CpGs was associated with increased GA, with increased or reduced birth size.

Among our top 20 hits from the A2 model, five CpGs were associated with gestational duration and four CpGs were associated with birth weight z-scores (Table 9) after Bonferroni adjustment. We produced forest plots of the associations for cg274402634, cg26843110, cg20340720 and cg17823829 (Figure 23) to demonstrate the pooled and cohort-specific associations with gestational duration and BW at those CpGs that yielded the strongest associations with MSDP. The primary hit, cg27492634, for which MSDP associated with lower DNA methylation, was not associated with GA or PTD, but was associated with BW (β = 0.92, P. Value = 6.71E-07) and BL (β = 1.07, P. Value = 4.21E-06) z-scores.

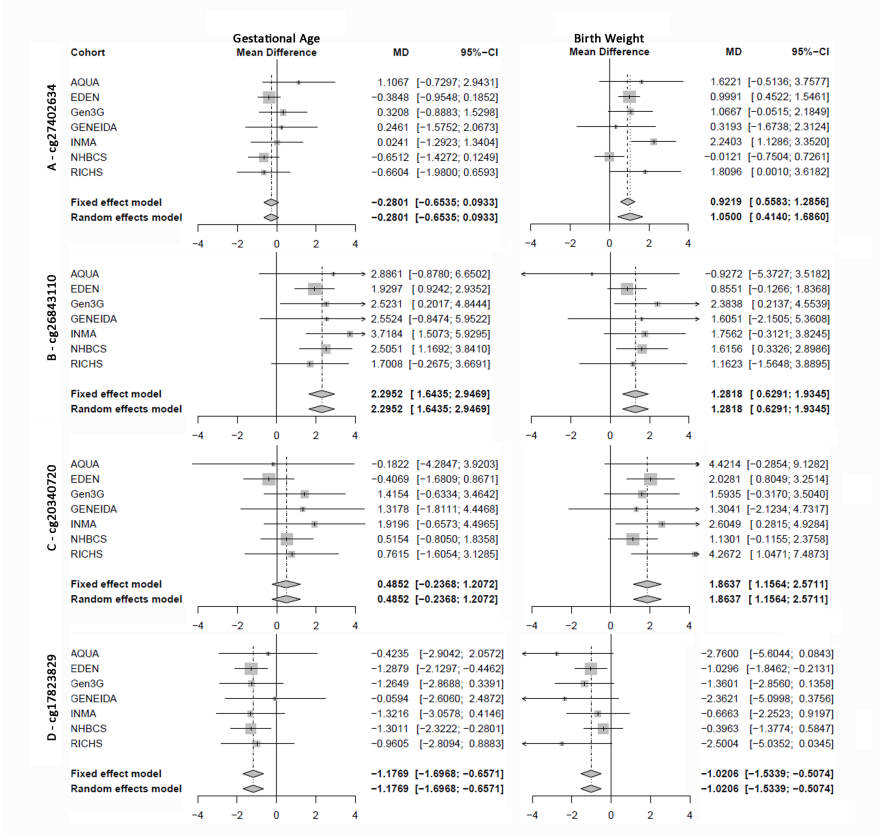

Figure 23. Forest plots of the pooled and cohort specific estimates of association between placental DNA methylation levels at (A) cg27402634, (B) cg26843110, (C) cg20340720, and (D) cg17823829 with gestational age and birth weight; models adjusted for maternal age, parity, maternal education, and putative cellular heterogeneity.

Figure 24 summarizes those hits that yielded BN-significance in both any- and sustained-MSDP, an eQTM relationship at 5% FDR, and an association with at least one of the reproductive outcomes at 5% FDR.

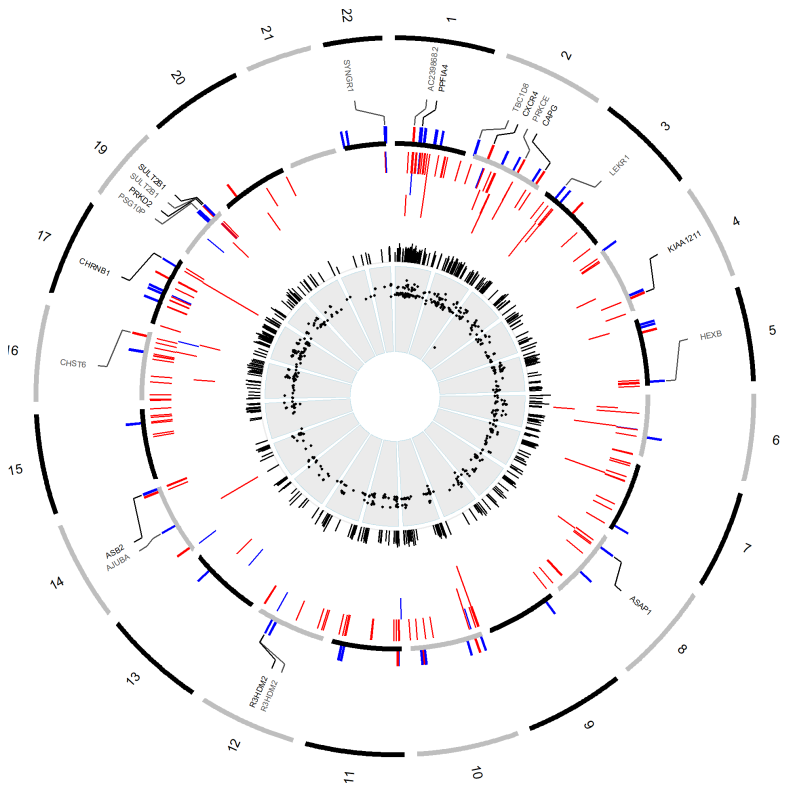

Figure 24. Circus plot summarizing all analyses for the 548 CpG sites that yielded Bonferroni-significant associations in both models A2 and S2. From inner to outer-band: Black points = differential methylation between any- and no MSDP; Black Bars = percent increase in magnitude of association from A2 to S2; Color Bars = Positive (red) or Inverse (blue) associations between DNA methylation and head circumference, birth length, birth weight, and gestational age, that yielded Bonferroni-significant associations; Color Bars = Positive (red) or Inverse (blue) associations between DNA methylation and nearby gene expression, that yielded association at an FDR < 0.05; Names of genes that yielded both an association between DNA methylation and expression with an FDR < 0.05 and an association between DNA methylation and at least one reproductive outcome after Bonferroni-adjustment.

3.2.8 Comparison with smoking-sensitive CpGs in cord blood

We assessed whether the epigenetic signatures of MSDP in the placenta were consistent with MSDP associations in cord blood. We compared the parameter estimated from our study that yielded associations with MSDP within 5% FDR to the parameter estimates of those CpGs described in cord blood also within a 5% FDR.[141] There was no overall correlation of the regression coefficients between the two tissues (Supplementary Figure S8). Only nine CpGs annotated to seven unique genes (*AHRR*, *CYP1A1, GNG12*, *PXN*, *RNF122*, *SLC23A2* and *ZBTB4* genes) yielded BN-significant associations in both tissues (Table 10). Of note, BN-significant CpGs in *CYP1A1* and *RNF122* genes showed opposite direction of association with MSDP in these two tissues.

Table 10. Consistently differentially methylated CpGs with MSDP in placenta and cord blood (from a published PACE meta-analysis)[141] that yielded Bonferroni-significant associations in both tissues.

			Cord Blood (S-MSDP*)		Placenta (S2-MSDP)	
CpG	**Gene**	**Chr**	**Effect**	**P. Value**	**Effect**	**P. Value**
cg25189904	*GNG12*	1	-0.024	1.38E-32	-0.020	2.54E-13
cg05575921	*AHRR*	5	-0.064	1.64E-193	-0.030	5.01E-09
cg21161138	*AHRR*	5	-0.022	1.78E-54	-0.026	5.25E-11
cg08327744**	*RNF122*	8	-0.010	1.37E-08	0.020	3.12E-09
cg15893360	*PXN*	12	-0.011	1.66E-08	-0.024	3.61E-08
cg12101586**	*CYP1A1*	15	0.045	3.67E-50	-0.045	7.15E-08
cg23680900**	*CYP1A1*	15	0.003	8.73E-08	-0.018	6.26E-21
cg07565956	*ZBTB4*	17	-0.006	2.50E-10	-0.023	4.02E-18
cg16547579	*SLC23A2*	20	-0.013	2.00E-10	-0.048	1.47E-08

4 EXPOSURE TO PRENATAL AND POSTNATAL TOBACCO SMOKING AND MOLECULAR SIGNATURES IN CHILDREN

4.1 Materials and methods

4.1.1 Study population

The Human Early Life Exposome (HELIX) study is a collaborative project across six established and ongoing longitudinal population-based birth cohort studies in Europe: the Born in Bradford (BiB) study in the UK,[233] the Étude des Déterminants pré et postnatals du développement et de la santé de l'Enfant (EDEN) study in France,[234] the INfancia y Medio Ambiente (INMA) cohort in Spain,[235] the Kaunus cohort (KANC) in Lithuania,[236] the Norwegian Mother and Child Cohort Study (MoBa)[237] and the Rhea Mother Child Cohort study in Crete, Greece[238] (Figure 25). The main aim of HELIX is to describe the early life exposome and its association with child molecular phenotypes and child health outcomes. HELIX used a multilevel study design with the entire study population totaling 31,472 mother-child pairs, recruited during pregnancy, in the six existing cohorts; a subcohort of 1301 mother-child pairs where biomarkers of environmental pollutants, omic signatures (blood methylation, transcriptomics including miRNAs, plasma proteins and serum and urinary metabolomics) and child health outcomes were measured at age 6-12 years; and repeat-sampling panel studies with around 150 children and 150 pregnant women aimed at collecting personal exposure data.

Cohort	Enrollment years	*n* subjects (total 32,000)
BiB—Born in Bradford, UK (Wright et al. 2013)	2007–2010	14,000
EDEN—Study of determinants of pre- and postnatal developmental, France (Drouillet et al. 2009)	2003–2006	2,000
INMA—Environment and Childhood, Spain (Guxens et al. 2012)	2004–2006	2,500
KANC—Kaunas Cohort, Lithuania (Grazuleviciene et al. 2009)	2007–2009	4,000
MoBa—The Norwegian Mother and Child Cohort Study, Norway (Oslo region) (Magnus et al. 2006)	1999–2009	8,000
Rhea—Mother–Child Cohort in Crete, Greece (Chatzi et al. 2009)	2007–2008	1,500

Figure 25. Participating birth cohorts in HELIX project

In the present study we selected children from the HELIX subcohort with complete data on main prenatal and postnatal variables of exposure to tobacco smoking and with at least one molecular layer (N=1203).

4.1.2 Exposure to tobacco smoking

4.1.2.1 Definitions of exposure to prenatal tobacco smoking

Information about maternal smoking habits during pregnancy was obtained from existing data collected in each cohort using non-harmonized questionnaires administered to the mothers, at least, on the first and third trimesters. Two variables for active maternal smoking during pregnancy were generated: i) any maternal smoking during pregnancy (any-MSDP: "yes" or "no"), and ii) sustained

maternal smoking, if the mother had smoked in the first and in the third trimesters (sustained-MSDP: "yes" or "no"). Mean number of cigarettes smoked per day during pregnancy was estimated averaging the mean number of cigarettes per day in the first and third trimesters.

Maternal exposure to secondhand smoking (mother-SHS) during pregnancy was assessed and harmonized in the ESCAPE project[239] and was defined as "yes" or "no"). The definition of passive or secondhand smoking was slightly different between cohorts: i) exposure at home, work or leisure places (INMA); ii) exposure at home or work (MoBa, Rhea, BiB); iii) exposure without specifying location (EDEN); and iv) partner smoking (KANC).

We combined all previous definitions to create a variable that captured both dose and duration of exposure to smoking in pregnancy. It had the following categories: "no exposure", "only mother-SHS exposure", "non-sustained smoker", "sustained smoker at low dose (=<9)" and "sustained smoker at high dose (>9)".

4.1.2.2 Definitions of postnatal exposure to secondhand smoking

Postnatal exposure to secondhand smoking (SHS) in children was assessed through a harmonized questionnaire administered to the parents as part of the HELIX project at ages 6-12 years. The questionnaire included questions about i) smoking at home by the mother, mother's partner or other people, and ii) attendance to indoor places where people smoke. From this information, we created a

variable of SHS with four levels: “no exposure”, “exposure to SHS only outside home”, “exposure to SHS only inside home” and “exposure to SHS inside and outside home”. This variable was reclassified in larger groups to assess: i) global exposure to SHS (global-SHS, “yes” or “no”) and ii) SHS only at home (home-SHS, “yes” or “no”).

Urinary cotinine in children was measured in HELIX using The Immulite® 2000 Nicotine Metabolite (Cotinine) 600 Test on an Immulite 2000 XPi from Siemens Healthineers at Fürst Medisinsk Laboratorium in Norway for which the Limit of Detection (LOD) was 3.03 μg/L. Due to low concentrations in children, a categorical variable was created categorizing urinary cotinine levels as detected or not detected considering a LOD of 3.03 μg/L.

4.1.3 DNA methylation

DNA was obtained from children’s peripheral blood (buffy coat) collected in EDTA tubes. DNA was extracted using Chemagen kit (Perkin Elmer) in batches of 12 samples in the same order of sample collection cohort by cohort. DNA concentration was determined in a Nanodrop 1000 UV-Vis Spectrophotometer (TheromoScientific) and also with Quant-iT™ PicoGreen® dsDNA Assay Kit (Life Technologies).

DNA methylation was assessed with the Infinium HumanMethylatio450 beadchip[240] from Illumina, following manufacturer's protocol at the University of Santiago de Compostela in the National Center of Genotyping (CeGen-USC) (Santiago de Complostela, Spain). 700 ng of DNA were bisulfite-converted using the EZ 96-DNA methylation kit following the manufacturer's standard protocol. After that, DNA methylation was measured using the Infinium protocol. In each plate, a HapMap sample was included and also 24 inter-plate duplicates of samples from the project. Samples were randomized taking into account sex and cohort.

After an initial inspection of the quality parameters of the data with the MethylAid package[241], probes with a call rate <95% based on a detection p-value of 1E-16 and samples with a call rate <98% were removed.[242] Methylation data was normalized using functional normalization with prior background correction using Noob.[211] ComBat applied to the data to correct the batch effect using the variable referred to slide (sets of 12 samples).[243]

Finally, some probes were filtered: control probes, probes to detect SNPs, probes to detect methylation in non-CpG sites, probes located in sexual chromosomes, cross hybridizing probes (probes that map to several genomic locations using the list described in Chen et al,[244] probes containing SNPs in any position of the sequence with a minor allele frequency (MAF) >5% and probes with a SNP at the CpG or at the single base extension (SBE) at any MAF (frequencies from the 1000 Genomes project considering several populations). CpGs were

annotated with the R package "IlluminaHumanMethylation450kanno.ilmn12.hg19"[220] obtaining information on gene, gene relative position and CpG islands. The number of initial and final samples and CpGs are shown in Table 11.

4.1.4 Gene expression

RNA was extracted from whole blood samples collected in Tempus tubes (Thermo Fisher Scientific) by using MagMAX for Stabilized Blood Tubes RNA Isolation Kit.[245] The quality of RNA was evaluated with a 2100 Bioanalyzer (Agilent) and the concentration with a NanoDrop 1000 UV-Vis Spectrophotometer. Samples classified as good RNA quality, which were subsequently analysed, had a RIN >5, a similar RNA integrity pattern in the visual inspection (Bioanalyzer) and a concentration >10 ng/ul. Mean values for the RIN, concentration (ng/ul), Nanodrop 260/280 ratio and Nanodrop 260/230 ratio were: 6.99, 107.36, 2.15 and 0.61.

Gene expression was assessed using the GeneChip® Human Transcriptome Array 2.0 (HTA 2.0) from Affymetrix at the University of Santiago de Compostela (USC) (Santiago de Compostela, Spain). Briefly, RNA samples were concentrated or evaporated in order to reach the required concentration (200 ng of total RNA were used as input). Amplified and biotinylated sense-strand DNA targets were generated from total RNA. Affymetrix Human Transcriptome Array 2.0 ST arrays were hybridized

according to Affymetrix recommendations using the Manual Target preparation for GeneChip Whole Transcript (WT) expression arrays and the Affymetrix labeling and hybridization kits.[246] Samples were randomized within each batch considering sex and cohort. Two different types of control RNA samples (HeLa and FirstChoice® Human Brain Reference RNA) were included in each batch, but they were hybridized only in the first's batches.

Raw data were extracted with the AGCC software (Affymetrix) and stored into CEL files and normalized with the GCCN (SST-RMA) algorithm at the gene level. Annotation to transcripts clusters (TCs) was done with the ExpressionConsole software using the HTA-2_0 Transcript Cluster Annotations Release na36. A transcript cluster is defined as a group of one or more probes covering a region of the genome reflecting all the exonic transcription evidence known for the region and corresponding to a known or putative gene.[247] After normalization several quality control checks were performed and four samples with discordant sex and two with low call rate were excluded. Control probes and probes in sexual chromosomes or probes without chromosome information were excluded. Probes with a DABG (Detected Above Background) p-value <0.05 were considered to have an expression level different from the background, and they were defined as detected. Probes with a call rate <1% were excluded from the analysis. Gene expression values were log2 transformed. Additional control of batch effect was done during the statistical analysis using a version of the surrogate variable analysis (SVA) method consistinting of the combination of two R

packages isva[248] and SmartSVA.[249] The number of initial and final samples and TCs are shown in Table 11.

4.1.5 miRNA

RNA was extracted from whole blood as has been detailed in the previous section and miRNA expression was determined using the SurePrint Human miRNA Microarray rel.21 from Agilent, which is a hybridization-based technique that contains probes for more than 2000 miRNA from miRbase v21.

MiRNA expression was assessed with the equipment and help from the Genomics Core facility at the Centre for Genomic Regulation (CRG) (Barcelona, Spain) and following Agilent's recommendations. After sample randomization by sex and cohort, samples were processed in batches of 24, which were hybridized in 3 different slides (8 samples per slide). A control sample, a total RNA mixture from 9 human tissues or cell lines (universal miRNA reference kit),[250] was included in 2/3 of the batches (a total of 39 control samples). 100 ng of total RNA were used as input, and RNA samples were concentrated or evaporated using a vacuum concentrator (SpeedVac) in order to reach the required RNA input. Fluorescently-labeled miRNA was obtained with the miRNA Complete Labeling and Hybridization kit, which was then hybridized onto the arrays according to Agilent recommendations. Raw data and GeneView files were extracted with the Feature extraction software (Agilent). miRNAs were annotated using a combination of

information from Agilent annotation ("Annotation_7056") and miRbase v21 (GRCh38) released in January 2017 ("annotation_miRBase_GRCh38_coordinates-gff3").[251] The coordinates from GRCh38 were annotated to hg19 reference genome.

A miRNA is considered as detected if its expression is different from the background or the standard error of its different probes is smaller than 3 times the expression signal. In HELIX samples, the average number of detected miRNAs was 420.43 (min: 71, max: 940) (Figure 26).

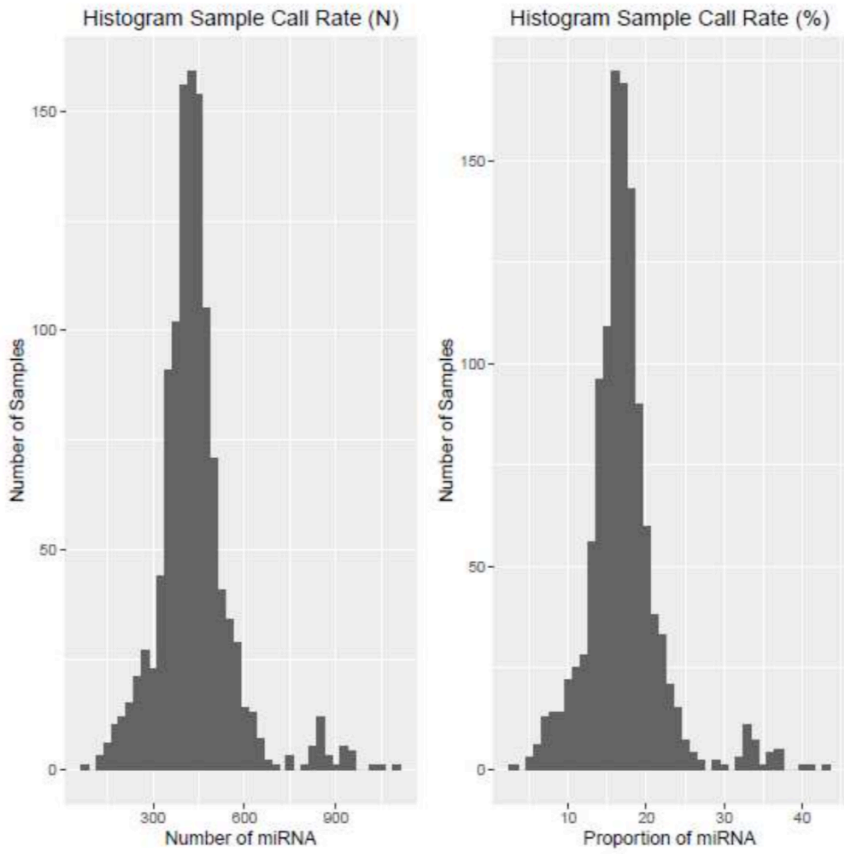

Figure 26. Histograms of sample call rate. The x-axis represents the number, or the proportion of detected miRNA and the y-axis represents the total number of samples.

None of the miRNA was detected in 100% of HELIX samples and the majority of miRNA were detected only in a small percentage of samples. 359 miRNAs were detected in 70% of the samples (Figure 27).

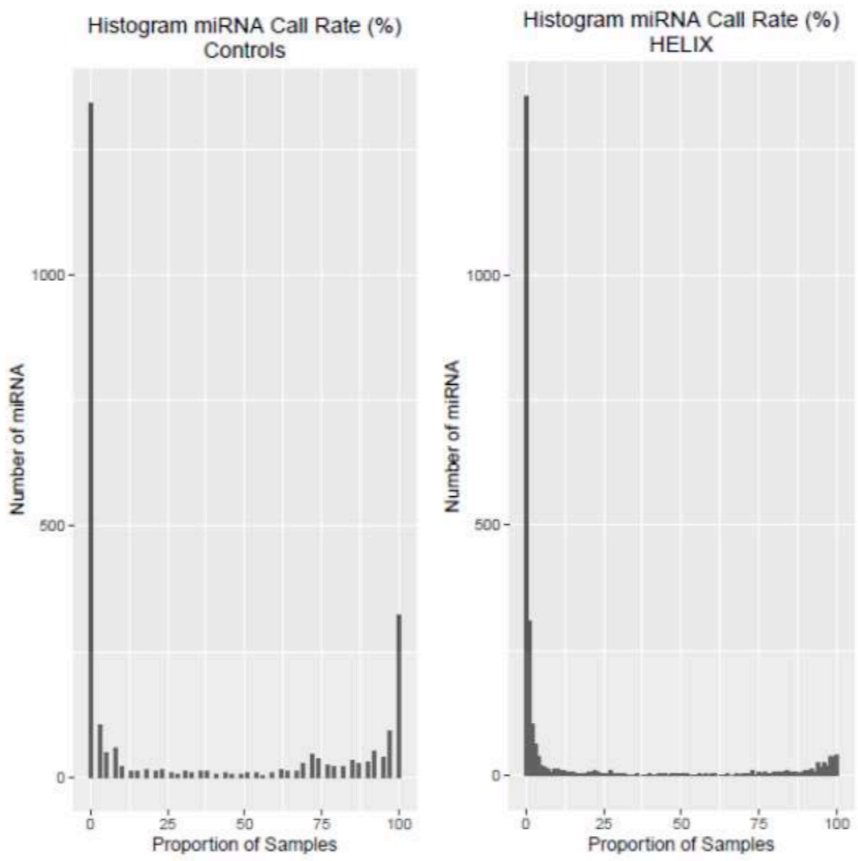

Figure 27. Histogram of miRNA call rate. The x-axis represents the proportion of samples and the y-axis represents the number of miRNAs detected at each percentage of samples.

After repeating 17 samples because they did not reach the laboratory quality control parameters, 6 samples were finally removed due to overall low quality. Microarray control probes, probes with a call rate <1% and miRNAs in sexual chromosomes were excluded from the analysis.

miRNA expression levels were normalized with the least variant set (LVS) method[252] with background correction using the Normexp

method in limma package.[253,254] LVS normalization method builds upon the fact that the data-driven housekeeping miRNAs that are the least variant across samples might be a good reference set for normalization. For the identification of housekeeping miRNAs a random sample of 50 HELIX samples was used. Final normalization using the selected miRNA reference set was done by applying the variance stabilization and calibration for microarray data (vsn) method. Additional control of batch effect was done during the analysis using the SVA method.[255,256] Final data was log2 transformed. The number of initial and final samples and miRNAs are shown in Table 11.

4.1.6 Proteome

Plasma protein levels were assessed through targeted mass spectrometry (MS) via Selective Reaction Monitoring (SRM) using the antibody-based multiplexed platform from Luminex.

A set of proteins were selected *a priori* based on the literature and on the Luminex kits commercially available from Life Technologies and Millipore. Three kits were selected assessing a total of 50 measurements with 43 unique proteins: Cytokines 30-plex (Catalog Number (CN): LHC6003M), Apoliprotein 5-plex (CN: LHP0001M) and Adipokine 15-plex (CN: LHC0017M). The proteins measured for each kit are in Supplemental Table S1.

Plasma samples from the HELIX children were analyzed at the CRG/UPF Proteomics Unit (Barcelona, Spain). All samples were randomized and blocked by cohort prior measurement to ensure a representation of each cohort in each plate (batch). First, samples were analyzed with the 30-plex kit, then with the 15-plex kit and finally, with the 5-plex kit, therefore undergoing three thaw/freezing cycles. For quantification, an 8-point calibration curve per plate was performed with protein standards provided in the Luminex kit and following procedures described by the vendor. Commercial heat inactivated, sterile-filtered plasma from human male AB plasma (Sigma CN. H3667) was used as constant controls to control for intra- and inter-plate variability. Four control samples were added per plate.

All samples, including controls, were diluted ½ for the 30-plex kit, ¼ for the 15-plex kit and 1/2500 for the 5-plex kit. Raw intensities obtained with the xMAP and Luminex system for each plasma sample were converted to ng/ml (5-plex kit) and to pg/ml (15 and 30-plex kits) using the calculated standard curves of each plate and accounting for the dilutions that were made prior measurement.

The percentages of coefficients of variation (CV%) for each protein by plate ranged from 3.42% to 36%. The limit of detection (LOD) and the limit of quantification (LOQ) were estimated by plate and then averaged. Several proteins had levels under the LOD or outside the linear quantification range (LOQ1-LOQ2). Only proteins with >30% of measurements in the linear range were kept in the database and the others were removed: IL-7, VEGF and GM-CSF had values

<LOD or <LOQ1 and lipocalin2, RANTES, resistin and SAA had values >LOQ2. Seven proteins were measured twice (in two different plex). Their correlations after log2 transformation ranged between 0.16 (IL-1Beta) to 0.56 (HGF). We kept the measure with higher quality and the other one was excluded from the analysis. In total, 14 proteins were excluded from the database.

Twenty samples were excluded due to having ten or more samples below the LOD or above the LOQ2. In addition, four more samples were excluded due to study design reasons or not having ethic consent.

Proteins that passed the quality control criteria mentioned above were log2 transformed and then the plate batch effect was corrected. After that, values below the LOQ1 and above the LOQ2 were imputed using a truncated normal distribution implemented in the truncdist R package.[257] The number of initial and final samples and proteins are shown in Table 11.

4.1.7 Serum metabolome

Serum metabololites was assessed using the targeted metabolomics Absolute*IDQ*™ p180 Kit (Biocrates Life Sciences AG), which allows the targeted analysis of 188 metabolites of different classes (amino acids, biogenic amines, acylcarnitines, glycerophospholipids, sphingolipids and sum of hexoses) and covering a wide range of metabolic pathways in one targeted assay.

The kit consists on a single sample processing procedure, although two separate mass spectrometry (MS) analytical runs, using a combination of liquid chromatography (LC) and flow-injection analysis (FIA) coupled to tandem mass spectrometry (MS/MS).

Isotope-labelled and chemically homologous internal standards were used for quantification. Of the total 188 metabolites measured, 42 metabolites were analyzed quantitatively by LC-ESI (Electrospray Ionization)-MS/MS with the use of external calibration standards in seven different concentrations and isotope labelled internal standards for most analytes. The other 146 metabolites were analyzed by FIA-ESI-MS/MS using a one-point internal standard calibration with 14 representative internal standards.

Serum metabolic profiles were acquired on a Sciex QTrap 6500 equipped with an Agilent 1100 series HPLC, at Imperial College of London (ICL) (London, UK), according to the manufacturer protocol.[258]

Samples were fully randomized and processed in 96-well plates. In every batch, it was included 4 blanks, 7 calibration standards and three sets of quality control samples (3 human plasma-based quality controls in three different concentrations, the NIST SRM 1950 and a commercially available serum QC material (CQC)).

Metabolites were quantified according to the manufacturer's protocol using Met*IDQ*™ Boron software for targeted metabolomic data

processing and management and results were shown in micromolar (μM) concentration units.

Blank PBS (phosphate-buffered saline) samples were used for the calculation of the LOD. The median value of the PBS samples on the plate was calculated as approximation of the background noise per metabolite and LOD was defined as three times this value (per metabolite and per batch).

From the 188 metabolites, six amines were excluded due to visual inspection of raw data showed no signal because of very low abundance. Of the 182 detected metabolites, we excluded from subsequent statistical analysis 5 metabolites with a % coefficient of variation (calculated from the 4 replicates of the SRM 1950) greater than 30% and the ones with a more than 30% of samples below the LOD. The final 177 metabolites are list in Supplemental Table S2. One sample, which was also very hemolyzed, was excluded after checking PCA.

Final data was log2 transformed. The number of initial and final samples and metabolites are shown in Table 11.

4.1.8 Urine metabolome

Urinary metabolic profiles were analyzed on a 14.1 Tesla (600MHz ^{1}H) NMR spectrometer at Imperial College London (ICL)

(London, UK). Samples were fully randomized to prevent potential bias and a pool from morning and night samples were used when it was possible.

Whilst the data acquisitions were untargeted, the data processing workflow enables to select the most abundant and prevalent metabolites for targeted data analysis and 44 metabolites were identified and quantified (Supplemental Table S3).

1D presat-NOESY spectra were imported into MATLAB (MathWorks, Massachusets, US) and were aligned to the pooled quality control sample using recursive segment-wise peak alignment (an algorithm based on cross-autocorrelation with a defined reference sample).[259]

For a subset of samples, poor peak alignment may result from multiple peak features occupying the same spectral region thus causing a lost in feature selectivity, and this is considered a major issue for NMR based metabolomics. Thus, it was also assessed the peak alignment for individual sample metabolite signals and computed Pearson's correlation coefficients R^2 between the aligned peak segment and the reference peak segment. While a high correlations coefficient represents good alignment, a low correlation coefficient represents that the sample peak segment has little resemblance to the reference peak segment. Finally, metabolite peak signals were estimated using trapezoidal numerical integration.

To validate the data processing workflow, metabolite signal data was compared to that processed using an alternative commercial software, "Chenomx",[260] which contains its own target compound library reference peak matching. Good correlations were found (Pearson R > 0.8) between these two different data processing methods for a large number of metabolites. For those metabolite signals whose identifies cannot be validated using Chenomx, their annotations were confirm using 2D NMR and/or STOCSY,[261] an approach that takes advantage of the multicollinearity of the intensity variables in the spectra dataset.

The sample concentration of a given metabolite can be estimated from the signal of the internal standard trimethylsilylpropanoic (TSP). Data was normalized using median fold change normalization which takes into account the distribution of relative levels from all 44 metabolites compared to the reference sample in determining the most probable dilution factor.

The coefficients of variation (CV) were assessed based on repeated analysis of 60 identical pooled quality control urine samples and these were analyzed at regular intervals during the sample run. The vast majority of metabolites/features achieved a coefficient of variation of <10%, indicating that sample run was of good analytical reproducibility.

The final dataset contains absolute concentration of 26 metabolites and semi-quantified concentrations of 18 metabolites. Final data was

log2 transformed. The number of initial and final samples and metabolites are shown in Table 11.

Table 11. Sample size and number of features included in each omic.

	Number of samples			Number of features		
	Initial	**After QC**	**With data on exposure to smoking**	**Initial**	**After QC**	**ENT**
Methylation	1200	1192	1105	485512	386518	-
Gene expression	1176	1158	958	64568	58254	-
miRNAs	961	955	895	>2000	1117	-
Proteome	1212	1188	1103	43	36	18.4
Serum metabolome	1209	1208	1128	188	177	60.3
Urine metabolome	1212	1211	1131	44	44	33.3

ENT: Effective Number of Tests[262]

Table 12 shows the number of children for which we got omics information (from 1 to 6 molecular layers). 834 children (69.3%) had data for all the omics; and 92.7% of them for at least 4 omics.

Table 12. Number and percentage of samples and the number of omics layers.

Number of omics	Number of samples	Percentage of samples	Cumulative percentage of samples
6	834	69.3%	69.3%
5	59	4.9%	74.2%
4	223	18.5%	92.7%
3	20	1.7%	94.4%
2	5	0.5%	94.9%
1	62	5.1%	100%

4.1.9 Statistical analyses

The correlation among the proportion of exposed children using different smoking definitions and different periods of exposure was calculate using the tetrachoric correlation test.[263]

To test the relationship between exposure to tobacco smoking, prenatal and postnatal, and molecular features assessed at age 6-12 years, we fitted linear regressions between each tobacco smoking variable and each molecular feature adjusting for covariates, using the limma R package,[254] implemented in omicRexposome.

The covariates were selected through directed acyclic graphs (DAGs) with the DAGitty tool.[264] We generated a DAG for the exposure to tobacco smoking during pregnancy (Supplementary Figure S9) and one for the postnatal exposure to SHS (Supplementary Figure S10). The following variables were selected as covariates in prenatal and postnatal models: i) cohort, ii) ethnicity based on existing data in the cohorts and defined in three categories (European ancestry, Pakistani or Asian and other, including African and native American among others), iii) maternal age as a continuous variable in years, and iv) maternal education, self-reported and categorized in low (primary school), medium (secondary school) and high (university degree or higher). In addition, prenatal models were adjusted for maternal body mass index (BMI) and postnatal models for child zBMI. Maternal BMI was calculated from pre-pregnancy self-reported weight and height and divided in four categories derived from World Health

Organization (WHO) definitions.[265] Child zBMI is a sex and age z-score calculated according to WHO reference curves[266,267] from anthropometric measurements obtained with a harmonized protocol in HELIX.

Since we wanted to assess the specific effect of postnatal smoking effects, we also included maternal smoking during pregnancy (A-MSDP or S-MSDP) as a covariate in the postnatal models. The prenatal models were run with and without adjusting for postnatal SHS (G-SHS or H-SHS) in order to capture direct and total (direct + indirect through postnatal SHS) effects.

Apart from covariates resulting from DAG diagrams, models were also adjusted for sex and child age, which were selected a priori by us due to study design criteria, as well as for specific covariates in models of each of the omics. Methylation models were adjusted for blood cell type composition (CD4+ and CD8+ T cells, natural killer cells (NK), monocytes, granulocytes and B cells), which was estimated using the Houseman algorithm[268] and the Reinius reference panel from raw methylation data.[269] Models of plasma proteins and serum metabolites were adjusted for time to last meal and hour of blood collection, and models for urine were adjusted for sample type (morning, night or the pool of both). Serum and urinary metabolites were additionally corrected for technical batch. To control for technical batch and blood cell type composition in transcriptomics data we used different versions of the SVA[248,256] method implemented in omicRexposome, as described in previous sections.

Missing data for covariates was imputed using a chained equations method[270] with the mice R package.[271] 20 imputed datasets were created,[270] although we only used the first imputation for our analyses. The proportion of missing data in the selected covariates were: 0.17% in maternal age, 0.75% in maternal BMI and 2.16% in maternal education.

The following models were tested:

(i) Any maternal smoking during pregnancy
Marker = Any-MSDP + covariates
Marker = Any-MSDP + covariates + post SHS (Global-SHS)
Marker = Any-MSDP + covariates + post SHS (Home-SHS)

(ii) Sustained maternal smoking during pregnancy
Marker = Sustained-MSDP + covariates
Marker = Sustained-MSDP + covariates + post SHS (Global-SHS)
Marker = Sustained-MSDP + covariates + post SHS (Home-SHS)

(iii) Global postnatal exposure to second hand smoking
Marker = Global-SHS + covariates + prenatal exposure (Any-MSDP)
Marker = Global-SHS + covariates + prenatal exposure (Sustained-MSDP)

(iv) Postnatal urinary cotinine levels
Marker = cotinine + covariates + prenatal exposure (Any-MSDPP)
Marker = cotinine + covariates + prenatal exposure (Sustained-MSDP)

Multiple testing correction was done in different ways. For methylation, gene expression and miRNAs we used the False Discovery Rate (FDR) - Benjamini Hochberg.[272] We also applied a more stringent correction calculating the Bonferroni threshold (1.29E-07 for methylation, 8.58E-07 for gene expression and 4.48E-

05 for miRNAs). For other omics we calculated the effective number of tests (ENT)[262] and divided the nominal P. Value by that number (18.4 for proteome, 60.3 for serum metabolome and 33.3 for utine metabolome). We considered as statistically significant any molecular feature that was significant in any of the two models tested for each smoking variable (e.g. biomarkers associated with any-MSDP after adjustment for global-SHS or for home-SHS).

The effect of smoking on DNA methylation was reported as a difference in methylation levels between exposed and unexposed children, while for the other omics, the effect is reported as a log2 fold change (log2FC). Finally, we repeated all the analysis restricting the sample to only children of European ancestry (N = 1083). In DNA methylation, to examine whether different definitions of exposure yielded increased or decreased magnitudes of association, we calculated the percent change in the coefficients between the two models $(|\beta_A| - |\beta_B|)/|\beta_A| * 100$.

All the analyses were done in R environment. The R packages omicRexposome,[273] MultiDataSet[274] and rexposome[275] were used to manage the omics and exposure data. Ggplot2,[276] qqman[277] and calibrate[278] R packages were used to create plots.

4.1.9.1 Expression quantitative trait methylation (eQTM) loci

We analyzed cis expression quantitative trait methylation loci (eQTM)[103] in 874 children with methylation and expression available

data. Cis effects were defined in a window of 1 Mb form the start of each transcript cluster (TC). The association between DNA methylation and gene expression levels was assessed via 1420 linear regressions in R.[254] Models were adjusted for cohort, child age, child zBMI, child ethnicity, sex and blood cell type composition. To control for gene expression batch effect, we estimated surrogate variables from the gene expression dataset protecting for cohort, child age, child zBMI, child ethnicity, sex and blood cell type composition with the SVA R package.[256] The effect of estimated surrogate variables on the gene expression was eliminated by getting the residuals and performing the regression analyses using them.

4.2 Results

4.2.1 Study population

Children included in the study had complete information on prenatal and postnatal exposure to tobacco smoking and at least one of the omics (N=1203). Table 13 shows the main characteristics of these children. Mean age was 8 years, there was 50% males, 10% of non-European ancestry, and 50% of the mothers with higher education.

Table 13. Characteristics of the population (N=1203)

	N/Mean (%/SD)
Cohort	
BiB	176 (14.6)
EDEN	171 (14.2)

INMA	215 (17.9)
KANC	189 (15.7)
MoBa	255 (21.2)
Rhea	197 (16.4)
Maternal age (years)	30.8 (4.8)
Maternal BMI	
<18.5	46 (3.8)
18.5-24.9	691 (57.4)
25-29.9	295 (24.5)
>=30	171 (14.2)
Maternal education	
Low	162 (13.5)
Medium	421 (35.0)
High	620 (51.5)
Child sex	
Female	547 (45.5)
Male	656 (54.5)
Child age (days)	8.1 (1.6)
Child zBMI	0.4 (1.2)
Child ethnicity	
European Ancestry	1083 (90.1)
Pakistani/Asian	93 (7.7)
Other (African, native American or other)	27 (2.2)

Child zBMI is a sex and age z-score calculated according to WHO reference curves.[266,267]

4.2.2 Exposure to tobacco smoking

Table 14 describes the proportion of children exposed to tobacco smoking, prenatally and postnatally.

Table 14. Proportion of children exposed to tobacco smoking, prenatally and postnatally.

Prenatal exposure:

	Any-MSDP	Sustained-MSDP
No	1027 (85.4%)	1027 (90.0%)
Yes	176 (14.6%)	114 (10.0%)

Postnatal exposure:

	Global-SHS	Cotinine
No	777 (64.6%)	993 (82.5%)
Yes	426 (35.4%)	210 (17.5%)

Prenatal and postnatal exposure:

Any- MSDP	**Postnatal SHS**			
	Global-SHS		**Cotinine**	
	No	**Yes**	**No**	**Yes**
No	710 (59.0%)	317 (26.3%)	918 (76.3%)	109 (9.1%)
Yes	67 (5.6%)	109 (9.1%)	75 (6.2%)	101 (8.4%)

14.6% of women smoked at some point during pregnancy and 10.0% smoked throughout all pregnancy. The proportion varied among cohorts, being BiB, EDEN, INMA and Rhea the ones with higher percentage of smoker mothers (~15% of sustained smokers) and KANC and MoBa the ones with the lowest percentage (<10%) (Figure 28).

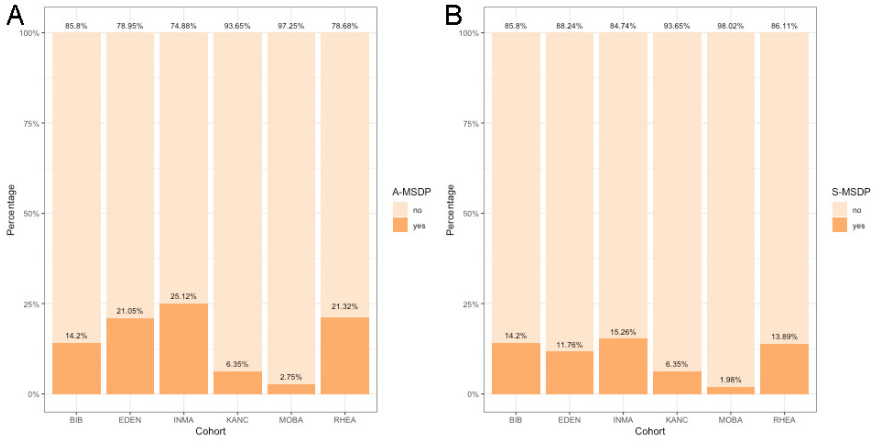

Figure 28. Percentages of any-MSDP (A) and sustained-MSDP (B), by cohort. A: Any; S: Sustained.

Figure 29 classifies women in function of their dose and duration of exposure to tobacco smoking in pregnancy (A, in all HELIX; and B,

by cohort). Around 50% of the population was not exposed to tobacco smoking during pregnancy, and 30% to only SHS. There were remarkable differences among cohorts. In MoBa, only 6.7% of the women were exposed to any sort of tobacco smoking; while in Rhea, the percentage raised to 90.9%.

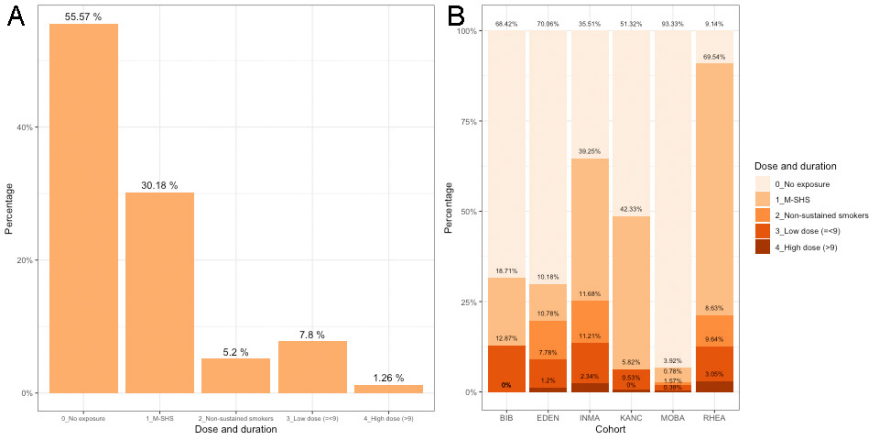

Figure 29. Percentages of dose and duration of MSDP in all the population (A) and by cohort (B). M: Mother

35.4% of children were postnatally exposed to global-SHS (Table 14). In all the cohorts, the percentage of exposed children was higher than 20% (Figure 30A), but KANC, Rhea and INMA had the highest values. 17% of the children had urinary cotinine levels over the LOD and were considered as exposed in our study (Table 14). The pattern of exposed children by cohort was similar using cotinine or questionnaire information, however the proportion of cotinine-exposed children was smaller in all the cohorts (Figure 30B). The correlation of SHS exposed children as defined from the questionnaire data and from the urinary cotinine measures was 0.7.

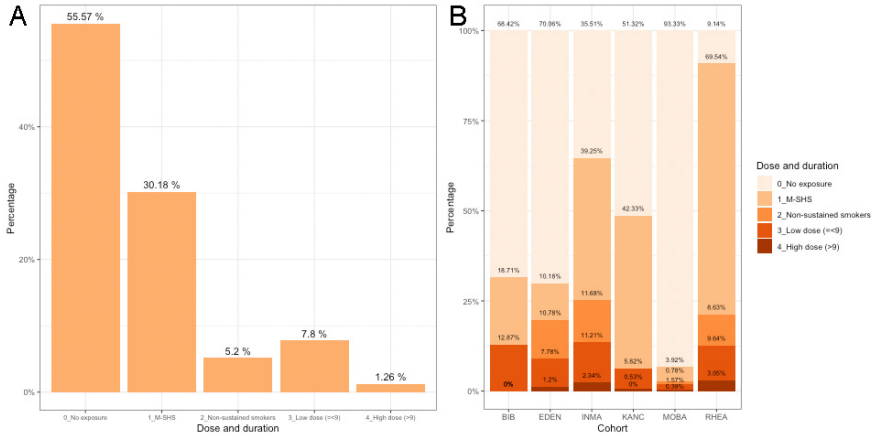

Figure 30. Percentage of children exposed to SHS by cohort assessed by parental-reported questionnaire (G-SHS) (A) and urine cotinine measurements (B)

When splitting the postnatal SHS in more detailed categories it was observed that around 17.2% of the children were exposed only outside home, 10.9% only at home and 7% in both places (Figure 31). Differences between cohorts were also found: in MoBa, children were only exposed outside home, whereas in the other cohorts some of them were also exposed inside home with different proportions; in Rhea, 25% of the children were exposed inside and outside home.

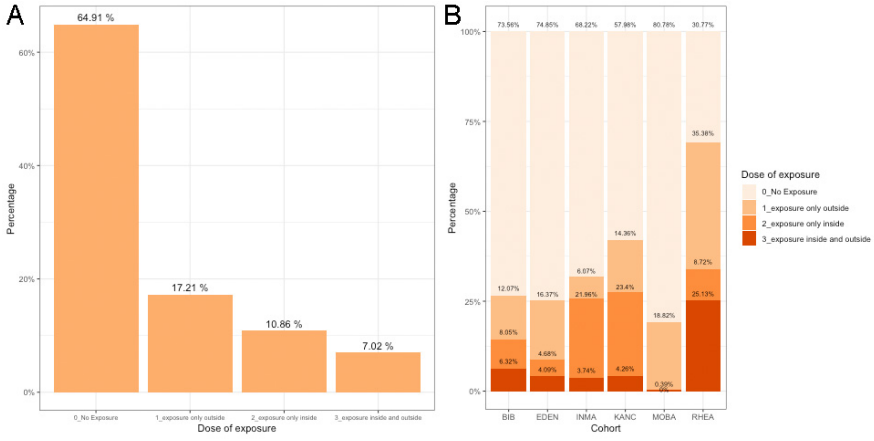

Figure 31. Percentage of dose of postnatal SHS in all population (A) and by cohort (B).

The correlations between any-MSDP and postnatal SHS were: 0.4 for global-SHS, ~0.6 for home-SHS; and 0.7 for cotinine. Similar correlations were observed for sustained-MSDP: 0.4 for global-SHS, 0.6 for home-SHS and 0.7 for cotinine. Looking at any-MSDP and global-SHS, ~9% of the children were exposed in both periods, 6% only prenatally, 26% only postnatally and 60% not exposed (Table 14).

4.2.3 Association between prenatal exposure to tobacco smoking and child molecular phenotypes

Any-MSDP and sustained-MSDP were associated with altered offspring methylation patterns at age 6-12 years. Table 15 shows the summary of the number of significant CpGs after multiple testing correction for the two smoking definitions (any and sustained) and the three different models (adjusted only for covariates and additionally adjusted for global-SHS or home-SHS).

Table 15. Number of significant CpG sites at the age of 6-12 years associated with MSDP.

MSPD	Adjustment	N	Lambda	N FDR CpGs	N BN CpGs
Any	Main covariates	1105	0.995	30	19
Any	Main covariates + postnatal G-SHS	1105	1.003	27	15
Any	Main covariates + postnatal H-SHS	1099	0.975	22	10
Sustained	Main covariates	1047	0.973	46	26
Sustained	Main covariates + postnatal G-SHS	1047	0.951	38	20
Sustained	Main covariates + postnatal H-SHS	1043	0.937	33	17

FDR: 5% False Discovery Rate; **BN:** Bonferroni correction (= 1.29E-07).

Lambda inflation factors ranged from 0.93 to 1.00 (Supplementary Figure S11) and did not substantially differ after adjustment for postnatal SHS variables.

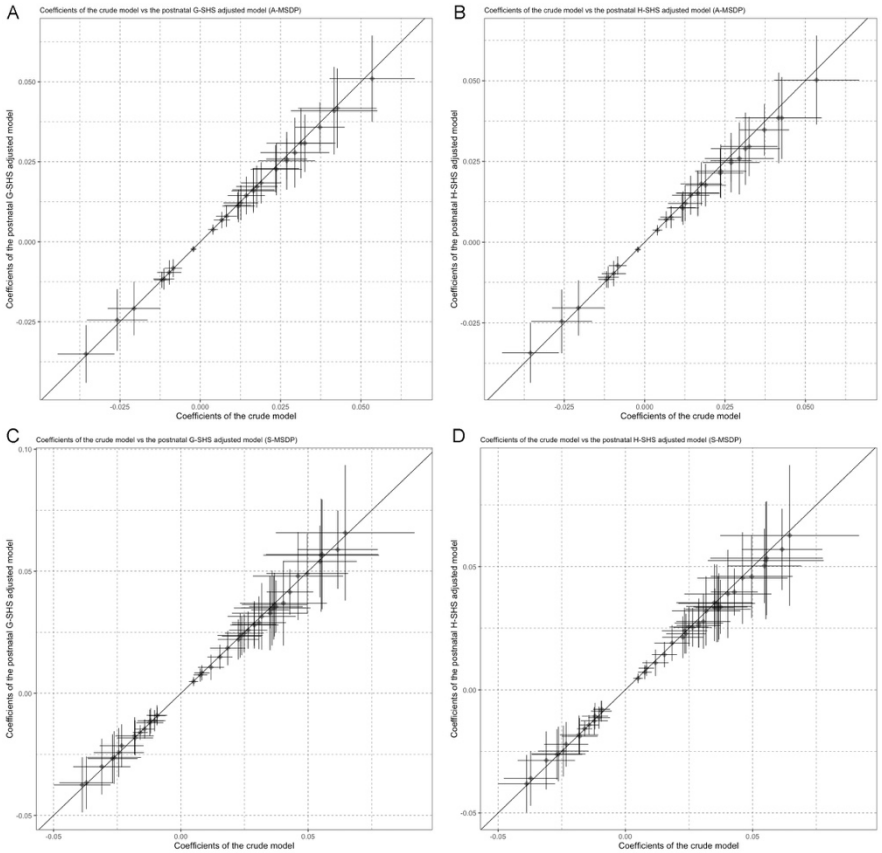

Figure 32. Comparison of the effects on DNA methylation between models adjusted only for main covariates (crude models) and postnatal SHS adjusted models: (A) any-MSDP and global-SHS; (B) sustained-MSDP and global-SHS; (C) any-MSDP and home-SHS (D) sustained-MSDP and home-SHS; x-axis: coefficients of the crude models; y-axis: coefficients of the postnatal SHS adjusted models.

The 52 CpGs differently methylated in at least one model at 5% FDR showed similar magnitudes of the effect in the models with and without adjustment for postnatal SHS (Figure 32), suggesting a direct prenatal effect, rather than an indirect effect through postnatal SHS.

In any-MSDP, the mean absolute change between methylation levels in models with and without adjustment for global-SHS was 4.0%, and 6.2% for home-SHS. In sustained-MSDP mean absolute differences were 3.2% for global-SHS and 5.5% for home-SHS

Given the similar results in the different models, only results of the models adjusted for SHS were commented on herein and we quoted as significant CpGs, those CpGs statistically significant after multiple testing in at least one of the models adjusted for postnatal SHS (either global-SHS or home-SHS). 5% FDR significant CpGs from models not adjusted for postnatal SHS are shown in Supplemental Excel Table S24.

Figures 33-34 shows the volcano and Manhattan plot of any-MSDP and sustained-MSDP after adjustment for global-SHS and home-SHS, respectively (Supplementary Figure S12: models non-adjusted for postnatal SHS). Around 30% of FDR-significant CpGs exhibit lower DNA methylation with any-MSDP and 32% with sustained-MSDP.

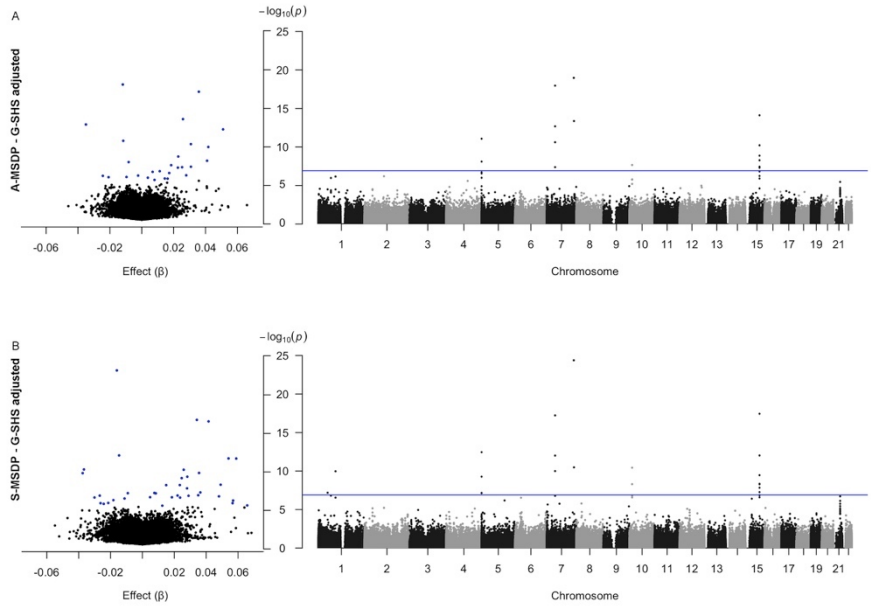

Figure 33. Volcano and Manhattan plots of the association between any-MSDP (A) and sustained-MSDP (B) and DNA methylation at 8 years adjusted for postnatal global-SHS. In the volcano plots, the x-axis shows the change in the DNA methylation (effect) and the y-axis the statistical significance (-log10 (P. Value)). 5% FDR-significant CpGs are shown in blue. Among them, 29.6 % (A) and 31.6 % (B) were hypomethylated. Sample sizes: N=1105 (A) and N=1047 (B). The Manhattan shows the -log10 (P. Value) in the y-axis and the genomic location of the CpGs in the chromosomes in the x-axis. The blue line represents the BN-significance threshold.

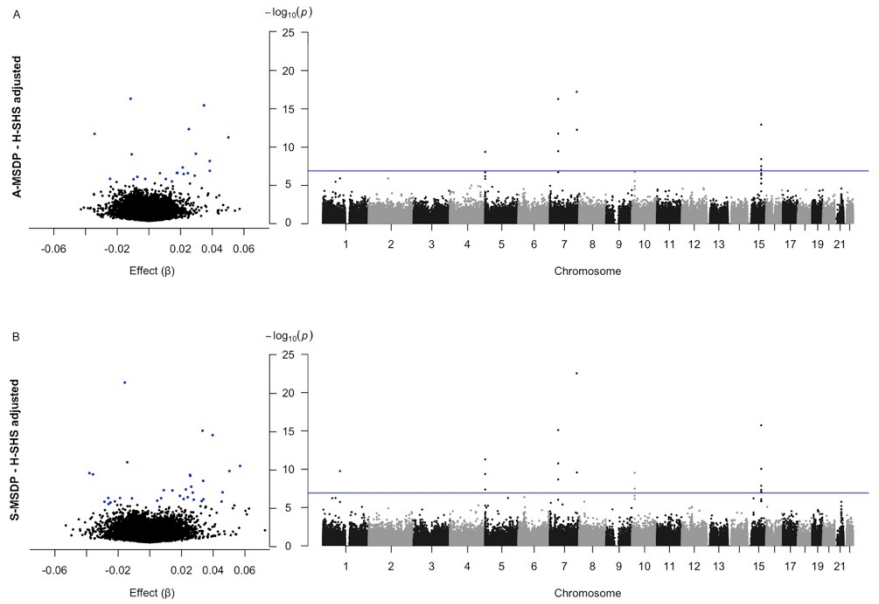

Figure 34. Volcano and Manhattan plots of the association between any-MSDP (A) and sustained-MSDP (B) and DNA methylation at 8 years adjusted for postnatal H-SHS. In the volcano plots, the x-axis shows the change in the DNA methylation (effect) and the y-axis the statistical significance (-log10 (P. Value)). 5% FDR-significant CpGs are shown in blue. Among them, 34.8% (A) and 33.5 % (B) were hypomethylated. Sample sizes: N=1099 (A) and N=1043 (B). The Manhattan shows the -log10 (P. Value) in the y-axis and the genomic location of the CpGs in the chromosomes in the x-axis. The blue line represents the BN-significance threshold.

At 5% FDR, a total of 41 unique CpGs were statistically significant in relation any-MSDP or sustained-MSDP: 25 shared among both smoking definitions, 3 only with any-MSDP and 13 only with sustained-MSDP, although all of them were at least nominally significant in both models (Supplemental Excel Table S22). These 41 CpGs were located in 18 different loci, defined as a region of <2 Mb, near these genes: *AGBL4* (number of significant CpGs: 1), *GNG12* (1), *GF1* (2), *NDST4* (1), *AHRR* (5), *PRRT* (1), *RADIL* (1), *MYO1G*

(4), *CNTNAP2* (2), *ZNF395* (1), *OLFM1* (1), *FRMD4A* (4), *FMN1* (1), *CYP1A1* (9) and *RUNX1* (4) and three intergenic regions (*2q12.*1 (1)*, 5q23.2* (1) and *7q11.22* (1)). Two loci were unique for A-MSDP (*2q12.1* and *NDST4*) whereas 8 loci were unique for S-MSDP (*AGBL4, 5q23.2, PRRT1, RADIL, 7q11.22, ZNF395, OLFM1, FMN1 and RUNX1*). In general, CpGs located in the same loci were altered in the same direction, except for the ones located near *AHRR* which exhibited both hypo- and hypermethylation. CpGs hypermethylated were separated around 450 kb from the hypomethylated ones. Table 16 shows a summary of the association for the 5% FDR CpGs, which in addition reached the BN threshold, and more detailed results are shown in Supplemental Excel Table S22.

Table 16. Association of BN significant CpGs that with any-MSDP or sustained-MSDP in either postnatal global-SHS or home-SHS adjusted models. Results for global-SHS adjusted model are shown ordered by position.

			Any-MSDP		Sustained-MSDP	
CpG	**Location**	**Gene**	**Effect**	**P. Value**	**Effect**	**P. Value**
cg18540492	1: 50489418	*AGBL4*	0.005	3.44E-05	0.008	6.58E-08
cg14179389	1:92947961	*GFI1*	-0.024	7.53E-07	-0.037	1.08E-10
cg17924476	5: 323794	*AHRR*	0.017	3.02E-07	0.025	5.39E-10
cg23067299	5: 323907	*AHRR*	0.007	2.08E-07	0.008	7.45E-08
cg11902777	5: 368843	*AHRR*	-0.008	8.59E-09	-0.009	7.07E-08
cg05575921	5: 373378	*AHRR*	-0.012	9.27E-12	-0.015	3.56E-13
cg19089201	7: 45002287	*MYO1G*	0.031	2.64E-11	0.036	1.02E-10
cg22132788	7: 45002486	*MYO1G*	0.025	4.51E-08	0.029	1.64E-07
cg04180046	7: 45002736	*MYO1G*	0.036	1.20E-18	0.042	6.04E-18
cg12803068	7: 45002919	*MYO1G*	0.051	2.28E-13	0.059	9.92E-13
cg25949550	7: 145814306	*CNTNAP2*	-0.012	1.17E-19	-0.016	4.26E-25
cg11207515	7: 146904205	*CNTNAP2*	-0.035	4.74E-14	-0.037	3.31E-11
cg25464840	10: 14372910	*FRMD4A*	0.016	1.88E-06	0.023	5.03E-09
cg15507334	10: 14372913	*FRMD4A*	0.018	2.37E-08	0.026	3.64E-11
cg00213123	15: 75019070	*CYP1A1*	0.011	1.68E-07	0.015	5.16E-09

cg05549655	15: 75019143	*CYP1A1*	0.026	8.34E-15	0.034	3.68E-18
cg17852385	15: 75019188	*CYP1A1*	0.023	5.28E-08	0.028	1.47E-08
cg13570656	15: 75019196	*CYP1A1*	0.042	6.67E-11	0.054	9.47E-13
cg12101586	15: 75019203	*CYP1A1*	0.031	3.97E-08	0.036	5.46E-08
cg22549041	15: 75019251	*CYP1A1*	0.041	5.84E-09	0.049	4.52E-09
cg11924019	15: 75019283	*CYP1A1*	0.023	1.47E-09	0.028	3.47E-10

BN = P. Value < 1.29E-07

We compared our list of significant CpGs with the 5% FDR CpGs from Joubert *et al*, the biggest study of maternal smoking during pregnancy and offspring DNA methylation to date.[138] Except for 6 CpGs [cg18540492 (*AGBL4*), cg01876548 (5q23.2), cg056767720 (*PRRT1*), cg07105221 (*RADIL*), cg17454592 (*FMN1*) and cg02869559 (*RUNX1*)], all the other 35 CpGs had already been described to be associated with sustained-MSDP in cord blood in the same direction of the effect as in our study. Moreover, 32 out of the 35 CpGs also remained altered at 7 years of age in the study of Joubert et al. (Supplemental Excel Tables S25-S27). Although 6 of our CpGs were not reported by Joubert *et. al.*, other CpGs in the same loci were found to be significant at 5% FDR in cord blood [*AGBL4* (number of significant CpGs: 2, in the same direction), 5q23.2 (6, 2 of thme in the same direction), *PRRT1* (8, in the same direction), *RADIL* (4, 3 of thme in the same direction) and *RUNX1* (8, in the same direction)]. *FMN1* (*Formin 1*) seems to be related to MSDP only in our study, however, other CpGs in the gene have been reported for current smoking although not in the same direction.[133]

The comparison of any-MSDP and sustained-MSDP showed stronger effects for sustained smoking (Figure 35). The mean

percentage of increased methylation from any-MSDP to sustained-MSDP models was 31.6% for models adjusted for global-SHS and 33.4% for models adjusted for home-SHS (Supplemental Excel Table S22).

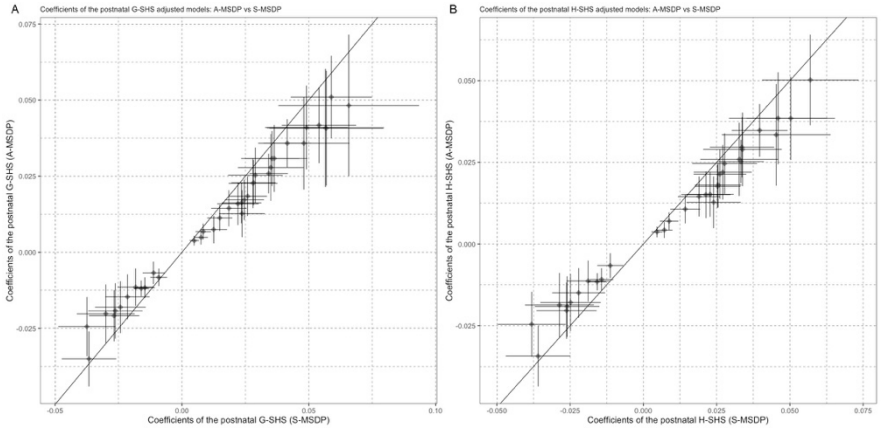

Figure 35. Comparison of the effects on DNA methylation between any-MSDP and sustained-MSDP in postnatal global-SHS (A) and home-SHS (B) adjusted models; x-axis: coefficients from sustained-MSDP model; y-axis: coefficients from any-MSDP.

The effect of dose and duration of the prenatal exposure to tobacco smoking on DNA methylation was investigated for the FDR-significant CpGs. At visual inspection, we identified 4 different patterns: i) CpGs exhibiting an increased (e.g. *MYO1G)* or decreased (e.g. *GNG12*, *RADIL* or cg05575921 (*AHRR*)) tendency with increasing MSDP dose and duration; ii) CpGs with an increased (e.g. cg17924476 (*AHRR*) or *OLFM1*)) or decreased (e.g. *GFI1*) tendency only with increasing dose of sustained-MSDP, but without response to any-MSDP*;* iii) CpGs which showed a saturated pattern when exposed to any-MSDP (e.g. cg11207515 (*CNTNAP2*), cg21161138

(*AHRR*) or *NDST4*); and iv) CpGs which showed a saturated pattern when exposed to sustained-MDP (e.g. *FRMD4A, PRRT1* or cg01664727 (*RUNX1*)). An example of each pattern is shown in Figure 36. The plots for all 41 FDR CpGs are shown in Supplementary Figures S13-S22. Some CpGs did not show a clear pattern in relation to dose and duration (e.g. *CYAP1A1*), and others in the same loci showed different patterns.

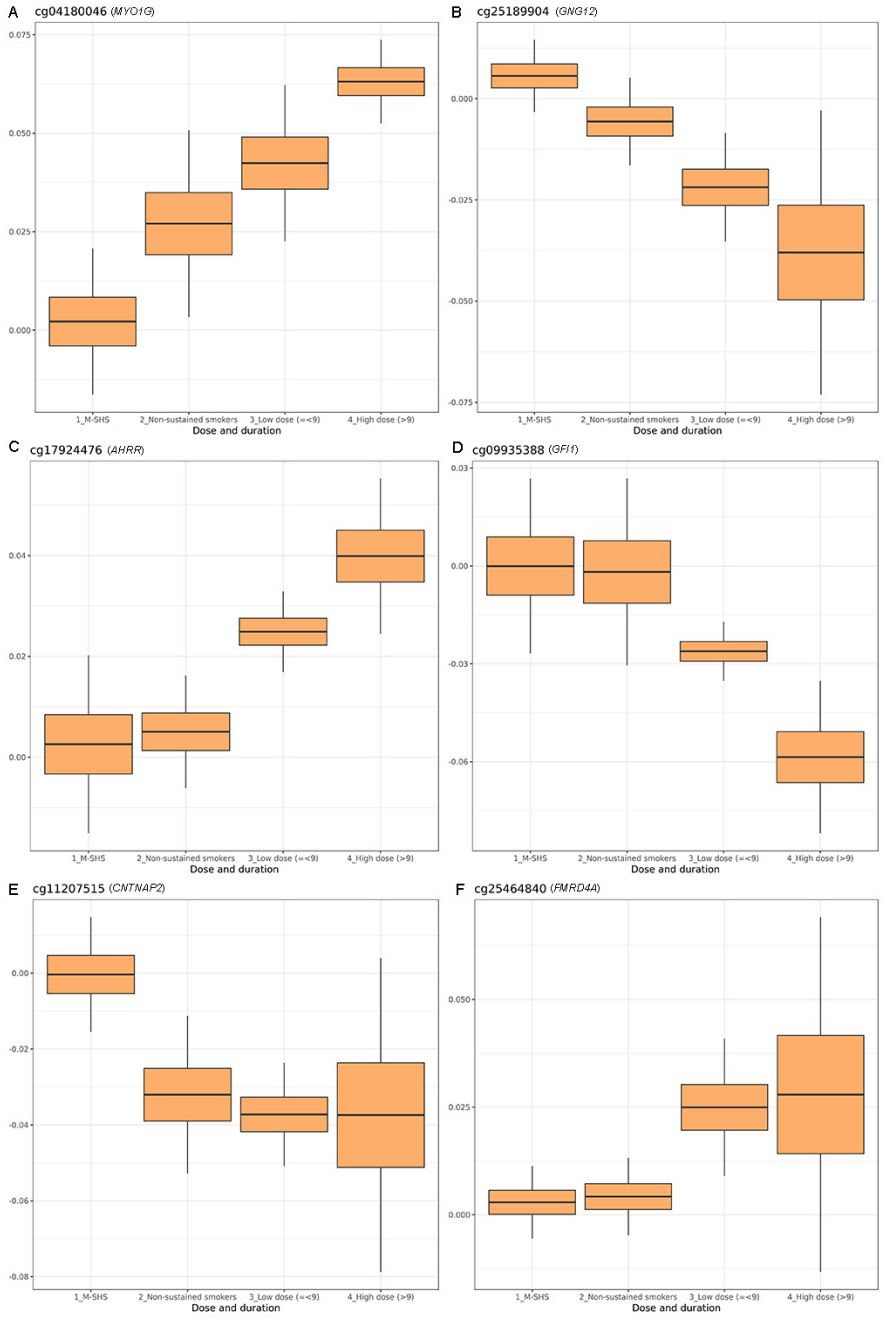

Figure 36. Box plots showing the change of methylation in relation to unexposed mothers for (A) cg04180046 (*MYO1G*), which shows an increased tendency; (B) cg25189904 (*GNG12*), which shows a decreased tendency; (C) cg17924476 (*AHRR*), which shows an increased tendency only in sustained-MSDP; (D) cg09935388 (*GFI1*), which shows a

decreased tendency only in sustained-MSDP; (E) cg11207515 (*CNTNAP2*), which shows a saturated pattern in any-MSDP; and (F) cg25464840, which shows a saturated pattern in sustained-MSDP.

We examined whether the 41 FDR-significant CpGs might be expression quantitative trait methylation (eQTM) loci. We identified 480 unique transcript clusters (TCs) with the start position within 500 kb upstream or downstream of the 41 candidate CpGs. At 5% FDR, we identified 15 CpG methylation to expression relationships (Supplemental Excel Table S28). This included 12 unique CpGs and 7 unique TCs, five of which showed positive associations (annotated to *ULK3* (number: 1)*, RUNX1-IT1* (2), *LOC285957* (2) and one located in 5p15.33), and two TCs showed inverse associations (annotated to *EXOC3* and *PNOC*). None of the 15 eQTMs genes corresponded to the genes annotated to the CpG sites with the Illumina annotation. All the significant methylation to expression relationships occurred between CpGs and genes located at <160 kb, except for one (>300 kb) (Figure 37).

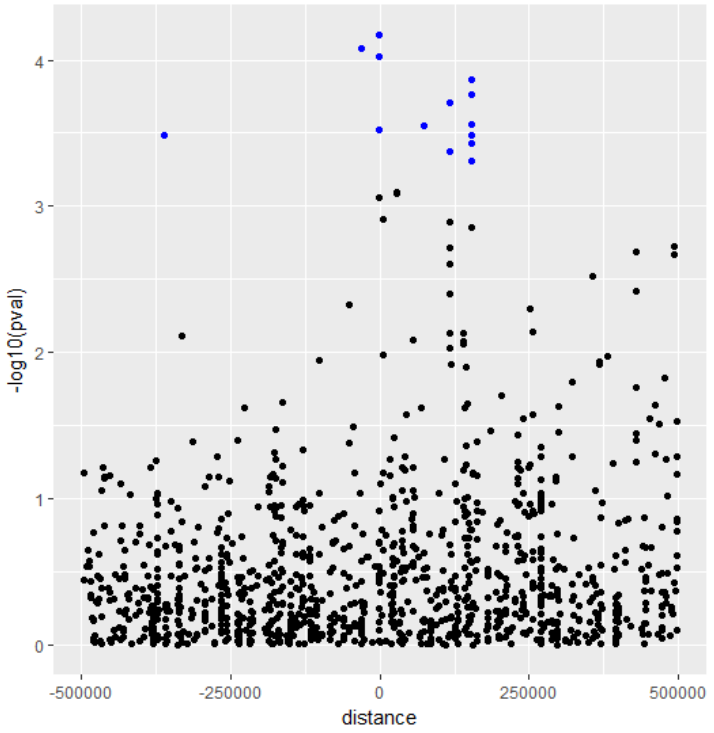

Figure 37. Graphic showing significance of methylation to expression relationships in relation to the distance between TC and CpG. In blue, associations at 5% FDR.

For the 15 FDR-associations, we observed that the smoking effect on DNA methylation, either hypo or hyper methylation, would result in upregulation of the gene expression (Table 17). An exception was found for the CpG-TC pair with the highest distance (>300kb) corresponding to TC05002792.hg.1-cg21161138 (*AHRR*).

Table 17. Summary of methylation to expression relationships at 5% FDR and the estimated effect of smoking on gene expression through DNA methylation regulation due to smoking.

Chr	CpG	CpG annot.	TC annot.	DNA methylation	CpG to TC effect	Estimated smoking effect on gene expression
5	cg11902777	*AHRR*	*EXOC3*	Hypo	-	Up
5	cg21161138	*AHRR*	5p15.33	Hypo	+	Down
8	cg17199018	*ZNF395*	*PNOC*	Hypo	-	Up
21	cg01664727	*RUNX1*	*RUNX1-IT1*	Hyper	+	Up
21	cg03142697	*RUNX1*	*RUNX1-IT1*	Hyper	+	Up
21	cg03142697	*RUNX1*	*RUNX1-IT1*	Hyper	+	Up
21	cg02869559	*RUNX1*	*RUNX1-IT1*	Hyper	+	Up
21	cg02869559	*RUNX1*	*RUNX1-IT1*	Hyper	+	Up
21	cg00994804	*RUNX1*	*RUNX1-IT1*	Hyper	+	Up
21	cg00994804	*RUNX1*	*RUNX1-IT1*	Hyper	+	Up
7	cg19089201	*MYO1G*	*LOC285957*	Hyper	+	Up
7	cg22132788	*MYO1G*	*LOC285957*	Hyper	+	Up
7	cg12803068	*MYO1G*	*LOC285957*	Hyper	+	Up
15	cg17852385	*CYP1A1*	*ULK3*	Hyper	+	Up
15	cg13570656	*CYP1A1*	*ULK3*	Hyper	+	Up

More information such as Transcript Cluster (TC), P. Value, complete annotation and distance between CpG and TC can be found in Supplementary Excel Table S28.

Moreover, 65 other nominal associations were detected, involving CpGs in 14 out of the 18 loci identified (Supplemental Excel Tables S22 and S28). The methylation level of CpGs in the gene body of *GFI1* and *AHRR* were positively associated with the expression of the same genes. *AHRR* showed hypo and hypermethylated CpGs in response to MSDP. Interestingly, they were located in two different regions of the gene (Figure 38) and they had a different effect on the expression of nearby genes (*AHRR, PDCD6* and *EXOC3*). MSDP-hypomethylated CpGs were positevely associated with expression, while hypermethylated inversely associated. This would result in

higher expression of *AHRR, PDCD6* and *EXOC3* in response to MDSP.

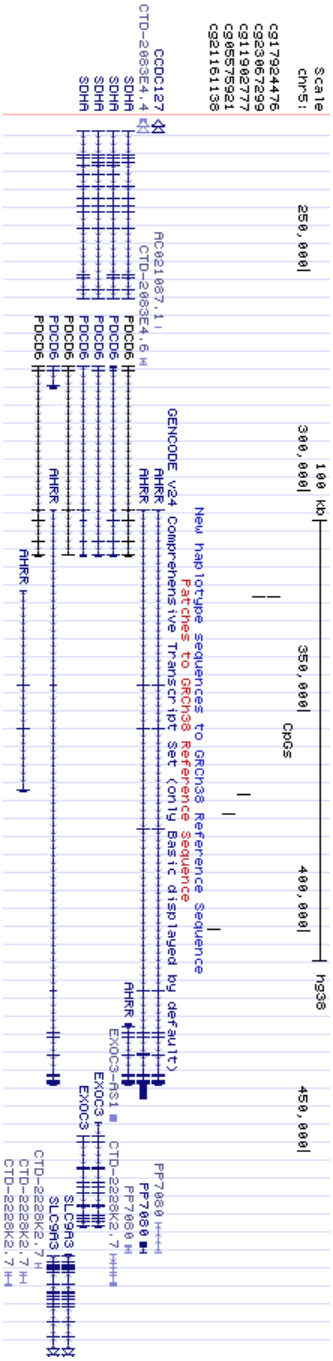

Figure 38. Location of the five CpGs in *AHRR* found to be associated with MSDP. Two of them, located in intron 1, were hypermethylated (cg17924476 and cg23067299) and three, located in intron 3 and 4, were hypomethylated (cg11902777, cg05575921 and cg21161138).

We, then, tested the association of MSDP and child gene expression at age 6-12 years. The transcription of none of genes identified thought the eQTM analysis was statistically significant in relation to MSDP (Table 18). At a nominal P. value, only the expression of *EXOC3* downregulated with MSDP. This effect would go in the opposite direction of the effect estimated from the eQTM analysis.

Table 18. Association of any-MSDP or sustained-MSDP with child blood gene expression adjusting for main covariates and global-SHS. The table shows the genes FDR-significant in the eQTM analysis, as well as those nominally significant for which the CpG annotated on the gene corresponded to the eQTM-gene. Genes are ordered by genomic position.

			Any-MSDP		Sustained-MSDP	
TC	**Position**	**Gene**	**Effect**	**P. Value**	**Effect**	**P. Value**
TC01002867.hg.1	1: 92940318-92952433	*GFI1*	-0.003	0.761	0.009	0.434
TC05000005.hg.1	5: 271736 - 438406	*AHRR; PDCD6*	-0.008	0.063	-0.005	0.326
TC05000006.hg.1	5: 443273 - 472052	*EXOC3*	-0.014	0.090	-0.024	0.014
TC05002792.hg.1	5: 26322 - 37720	-	-0.035	0.065	-0.036	0.112
TC07002924.hg.1	7: 44998599 - 45000392	*LOC285957*	-0.017	0.606	-0.023	0.547
TC08000217.hg.1	8: 28174503 - 28200872	*PNOC*	-0.001	0.950	-0.008	0.611
TC15001649.hg.1	15: 75128459 - 75135552	*ULK3*	-0.006	0.437	-0.001	0.928
TC21000425.hg.1	21: 36410233 - 36411723	*RUNX1-IT1*	0.033	0.207	0.057	0.063
TC21000946.hg.1	21: 36410233 - 36411723	*RUNX1-IT1*	0.019	0.466	0.028	0.358

The sensitivity analysis in children of European ancestry gave attenuated P. Values and 6 out of the 41 CpGs did not reach 5% FDR, likely due to a smaller samples size, but effect sizes remained of similar magnitude (Supplemental Excel Tables S22-S23; Figure 39). Mean difference in DNA methylation between these models was 3.2% for any-MSDP and 1.4% for sustained-MSDP.

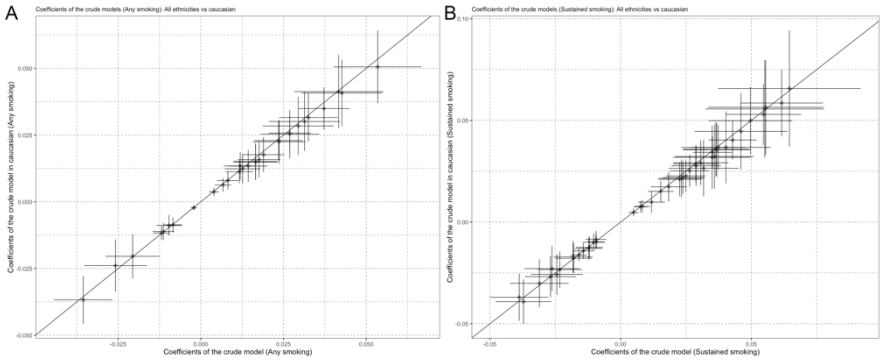

Figure 39. Comparison of effects on DNA methylation between models including all the population and models including only children of European ancestry in (A) any-MSDP and (B) sustained-MSDP, both adjusted for global-SHS; x-axis: coefficients from crude model including all the population; y-axis: coefficients from crude model including only children of European ancestry

Besides DNA methylation, few associations were found between prenatal exposure to smoking and other molecular layers (Supplemental Excel Tables S29-S30). The *Teratocarcinoma-derived growth factor 1 (TDGF1)* gene (TC3003404.hg.1) showed lower expression levels in children exposed to any-MSDP in home-SHS adjusted model (effect = -0.071, P. Value = 6.58E-07). On the other hand, alanine and lactate were increased in children born from sustained smokers (effect = 0.189 and 0.174, P. Value = 3.93E-04 and 5.19E-04, respectively) and urea and acetone decreased only in

European ancestry population (effect = -0.137 and -0.443, P. Value = 8.04E-04 and 1.02E-03, respectively). We also investigated dose and duration effect response, but no clear patterns were detected (Supplementary Figure S23).

4.2.4 Association between postnatal exposure to secondhand smoking and child molecular phenotypes

In contrast to prenatal exposure to tobacco smoking, postnatal SHS was related to protein and metabolite levels, but not to DNA methylation (Supplemental Excel Tables S31-S32).

Higher levels of PAI1 (plasminogen activator inhibitor 1) were found in children with detectable cotinine levels (effect = 0.316, P. Value = 8.95E-04). PAI1 levels followed an increased tendency with increasing frequency of exposure, being children exposed both inside and outside home the ones with the highest levels (Figure 40).

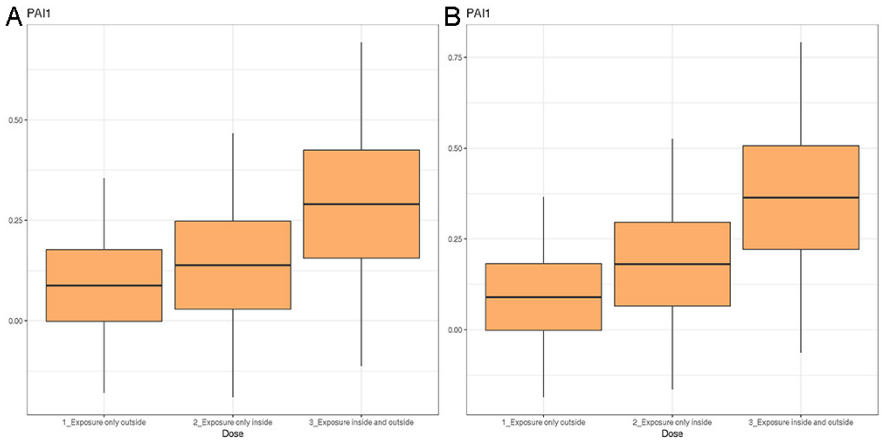

Figure 40. Boxplots showing the change of PAI1 levels in relation to unexposed children. (A) Adjusted for any-MSDP; (B) Adjusted for sustained-MSDP.

On the other hand, global-SHS was related to lower levels of serum sphingolipid SM-(OH)-C16.1 (effect = -0.065, P.Value= 3.48E-04) and acilcarnitine C9 (effect = -0.048, P. Value = 1.37E-03) , while cotinine was associated with lower levels of serum acyl-alkyl-phosphatidylcholine C38:0 (PC ae C38.0) (effect = -0.102, P. Value =) and urea (effect = -0.112, P. Value = 8.18E-04). Results were consistent when restricting the analysis to children of European ancestry. Furthermore, in the only European ancestry sample, we detected additional statistically significant associations: tyrosine with global-SHS (effect= -0.108, P. Value = 6.29E-03) and lower levels of 7 phospholipids (2 unsaturated diacyl(aa)-PCs and 5 acyl-alkyl(ae)-PCs) with cotinine levels (max effect = -0.127, min P. Value=1.44E-04).

We, then, took a closer look at these biomarkers by exploring its association with SHS outside, inside and both inside and outside home (Supplementary Figure S24). For serum metabolites (PC aa C38:0, C9 and SM (OH) C16:1) we did not detect a clear pattern of dose-effect response; whereas for urea we detected lower levels in those children exposed at home or at both places.

5 DISCUSSION

5.1 Proportion of children exposed to tobacco smoking

Despite the efforts of public health campaigns to warn of the dangers of tobacco smoking and to promote its cessation, there are still mothers that smoke during pregnancy, or children which are exposed to SHS.

In our studies, including a total of 11 cohorts around the world, the percentage of women who smoked at any point during pregnancy is ~20%, and ~10% of them smoked in a sustained way throughout pregnancy, for both groups (cohorts from HELIX Project and cohorts from PACE Consortium). However, there are big differences between cohorts ranging from 6% to 31% in PACE cohorts and from 2% to 25% in HELIX cohorts.

For postnatal SHS exposure, including 6 different European cohorts, ~35% of children were exposed based on questionnaire data, which reflects a more continued exposure, ranging from 19% to 70% depending on the cohort. However, ~17%, ranging from 1% to 39% of children were exposed based on detection of urinary cotinine, which reflects recent exposure. Correlations between prenatal and postnatal exposures ranged from 0.4 to 0.7, being cotinine the measure with higher correlation with MSDP. Cotinine reflects recent exposures, which probably corresponded to parents smoking.

5.2 Maternal tobacco smoking during pregnancy and placental DNA methylation, and its link with reproductive outcomes

5.2.1 Placental DNA methylation signatures of maternal tobacco smoking during pregnancy

We performed meta-analyses of differential DNA methylation associated with any-MSDP among seven cohorts and sustained-MSDP among three cohorts, identifying 1224 CpG sites that were associated with either of these smoking exposures in placenta.

The top MSDP-loci in placental tissue, cg27402634 which is upstream of *LEKR1 (Leucine, glutamate and lysine rich 1)*, was identified in prior EWAS of placental tissue,[151] but not in prior EWAS of peripheral or cord blood, and thus appears to be a placenta-specific response to MSDP. DNA methylation at cg27402634 yielded dramatically lower DNA methylation levels in associations with exposure, much larger in magnitude (~25% difference for sustained MSDP) than has observed in most exposure-focused EWAS, though comparable to a CpG site in *AHRR* from prior EWAS of current smoking and blood methylation.[133] Additionally, placental DNA methylation at cg27402634 was inversely correlated with the expression of *LEKR1* and we found that higher methylation at cg27402634 was associated with larger BW and BL z-scores. However, the consequences of perturbations to the epigenetic

regulation of cg27402634 or the expression of *LEKR1* in human placental tissues are not well understood.

Despite limited understanding of the biological roles of *LEKR1*, GWAS findings provide evidence that genomic features within this region appear to be involved in fetal growth and possibly metabolic programming. For instance, rs13322435 or its proxy (rs900400) have been associated with different parameters of fetal growth,[279] adiposity in newborns,[280281] maternal adiponectin levels in pregnant women and cord blood leptin (BW increasing allele – increased levels of leptin and lower of adiponectin),[281] and insulin release after an oral glucose challenge (BW lowering allele – increased insulin release),[282] while rs900400 BW-increasing allele is associated with higher expression of *TIRAP (TCDD Inducible Poly(ADP-Ribose) Polymerase*) in adipocytes.[281] Together with these previous genetic association reports and our current study, it seems that this locus on chromosome 3 contains very active determinants of growth regulation and metabolic activity, and that placental DNA methylation here is highly responsive to maternal smoking. Future work is necessary to investigate whether the placental epigenetic regulation at this locus specifically influence overall fetal growth and metabolic functions of the developing fetus.

We identified numerous other notable MSDP-associated loci in addition to those cg27402634 and *LEKR1*. The CpG with the second smallest P. Value, cg20340720, located within *WBP1L* (*WW Domain Binding Protein 1 Like*, also annotated as *C10orf26*), was also

positively associated with BW and BL z-scores. Genetic variants nearby to this CpG have been related to BW[283] and to blood pressure.[284] Methylation at the third most statistically significant hit from our analysis, cg26843110 which is within the body of the *EDC3* (*Enhancer Of MRNA Decapping 3*) gene, was associated with increased GA, inversely associated with the expression of *UBL7-AS1*, a non-coding RNA, and positively associated with the expression of *CSK* (*Carboxyl-terminal Src kinase*) expression which is involved in fetal growth and development through trophoblast cells.[285] However, the DNA methylation associations with the expression of *UBL7-AS1* and *CSK* were only nominally significant. Finally, cg17823829 produced the second largest effect size among our most statistically significant hits. This CpG is within *KDM5B (Lysine Demethylase 5B)*, and its methylation was inversely associated with the expression of *PPFIA4* (*PTPRF interacting protein alpha 4*) gene, which has been shown to be induced in response to hypoxia.[286]

Though we highlight the associations involving MSDP, DNA methylation, gene expression, BW and GA at the four CpG sites that yielded the strongest magnitudes of association with MSDP, we must also emphasize that many more CpGs were differently methylated, giving a specific epigenetic signature. Our enrichment analyses identified numerous pathways that are critical to placental growth and development, such as vascularization, hormone signaling and inflammatory cytokines. Multiple pathways involved vascular endothelial factors (VEGFs) and nerve growth factors (NFG). The

VEGFs and their receptors are required for all steps of placental vascularization.[287] NGF regulates the growth and maintenance of neuronal cells, and in placental tissues NGF modulates immune activity, inflammation and angiogenesis.[288] Thus, our findings may be related to perturbed placental vascularization or angiogenic signaling. Placental blood flow and vascular development are essential for normal placental function and critical for fetal growth and development. Altered placental vasculature is the most common placental pathology identified in pregnancy complications,[287] and MSDP has been found to decrease placenta vascularization.[289,290] Our CpGs were also enriched for genes regulated by specific transcription factors, most notably GATA1 and GATA2. Together with RUNX1, PPARG and TP63, GATA factors are part of the core transcriptional regulatory circuit needed for trophoblast differentiation and their expression is maintained in mature placenta.[287,291] Placentas lacking *PPARG* have lethal defects in placental vascularization,[284] and angiogenic activity is reduced in placentas lacking *GATA2*.[292] Additionally, our MSDP-associated CpGs were enriched in placental enhancers as defined in ROADMAP, while depleted in transcription start sites (TSS), inactive states, PMDs[112] and DMRs[222], overall suggesting that our hits lie in active regulatory regions in placental tissue.

Given the growing recognition that altered placental function can have severe consequences on pregnancy outcomes and fetal development,[293] our study provides important insights into the

placental epigenetic responses associated with a relatively common and highly toxic exposure, prenatal tobacco smoke.

5.2.2 Smoking-associated placental DNA methylation signatures and offspring health outcomes

The methylation of 341 of the 1224 smoking-associated CpG was related to at least one birth outcome (GA, BW, BL, or HC z-scores) in our cohorts. 85% of the CpGs were associated exclusively with GA, 12% with birth size and 3% with GA and birth size, even though they were standardized for GA, suggesting independent associations with both gestational duration and fetal growth. Among the 341 CpGs, the ones with positive associations with reproductive outcomes tended to be hypomethylated with MSDP, while the ones with inverse associations tended to be hypermethylated.

Phenotype enrichment analyses of the list of genes annotated to our set of MSDP-associated CpGs, highlighted a number of conditions that are part of the metabolic syndrome, which appears in adulthood and for which some links with MSDP have been reported,[294,295] as well as asthma and respiratory function.

Furthermore, 21 of the 74 SNPs that have previously associated with birth size or gestational duration,[227–229,283,296] were in similar genomic proximity (within 1 Mb) to our MSDP-associated CpGs. It is also noteworthy that almost all autosomal genetic loci of birth outcomes were within at least 1 Mb of the CpGs that yielded at least 5% FDR

with any or sustained MSDP. Future research is needed to characterize the consequences of this convergence of genetic effects on growth and epigenetic responsiveness to MSDP within similar proximities. It is possible that MSDP, genetics and epigenetic in these regions yield additive or interactive effects on birth outcomes. Previous studies have addressed this question in blood.[297,298] Importantly, methylation probes that contained SNPs were eliminated from our study.

Overall, our findings suggest that some of these placental epigenetic variations may be intermediate molecular markers between MSDP and reproductive outcomes, though it is still unclear if these epigenetic variations are causal mediators within these relationships or reflecting other processes.

5.3 Exposure to prenatal and postnatal tobacco smoking and molecular signatures in children

We performed association analyses between any- and sustained-MSDP as well as maternal and child SHS and different molecular layers in children.

5.3.1 Exposure to prenatal tobacco smoking and child molecular phenotypes

Blood DNA methylation was the only molecular layer measured in childhood which was associated with *in utero* tobacco smoking, suggesting persistent epigenetic imprints, which were not observed for transcriptomics, miRNA expression, proteins or metabolites. Of note, methylation-MDSP models were adjusted for postnatal exposure to SHS, and therefore eliminating potential direct cross-sectional postnatal effects.

Exposure to prenatal tobacco smoking was associated with DNA methylation of 41 CpG sites located in 18 different loci corresponding to *AGBL4*, *GNG12*, *GF1*, *NDST4*, *AHRR*, *PRRT*, *RADIL*, *MYO1G*, *CNTNAP2*, *ZNF395*, *OLFM1*, *FRMD4A*, *FMN1*, *CYP1A1* and *RUNX1* genes and three intergenic regions (*2q12.*1*, 5q23.2* and *7q11.22*). All loci had previously been reported of the same direction of the effect in relation to MDSP at birth and at older ages,[133,299] and their magnitudes of the effect were similar to ours (from -0.04 to 0.07). However, none of the previous studies had incorporated substantial transcriptomics data from the same subjects for the biological interpretation of the findings, and only a few of them have analyzed dose and duration of the exposure.

Interesting findings are the ones located in *AHRR* (*Aryl-Hydrocarbon Receptor Repressor*). We identified three hypomethylated (in intron 3 and 4: cg11902777, cg05575921 and 21161138) and two hypermethylated (in intron 1: cg23067299 and cg17924476) CpGs in response to MSDP. cg05575921 has been largely reported to be associated with current[133], former[133] and MSDP,[299] and has been

found to be hypomethylated in different tissues, including adult blood,[133] cord blood,[299] adult adipose tissue[300] and placenta.[299,301] Moreover, we also found that cg05575921 is an eQTM for *AHRR* in peripheral blood and in placenta, suggesting upregulation of this gene in response to tobacco. Interestingly, the cg23067299 in intron 1 was only reported for MSDP and not for current smoking,[301] whereas cg17924476 was not reported in newborn tissues, but in older offspring.[145]

Looking at eQTMs and taking into account methylation changes due to MSDP, we observed that CpGs located in the body of *AHRR,* either hypo or hypermethylated, would result in up-regulation of *AHRR, PDCD6* or *EXOC3*. *PDCD6 (Programmed Cell Death 6*) and *EXOC3 (Exocyst Complex Component 3)* are the proximal genes downstream and upstream *AHRR*. The role of these genes in tobacco response is unknown. *PDCD6* is a calcium sensor involved in endoplasmic reticulum (ER)-Golgi vesicular transport, endosomal biogenesis or membrane repair; and *EXOC3* is a component of the exocyst complex involved in the docking of exocytic vesicles with fusion sites on the plasma membrane.

AHRR is a negative regulator of AhR (Aryl Hydrocarbon Receptor) activity, which is activated in response to environmental pollutants and components of cigarette smoke.[300] Once AhR is activated by an agonist, it is translocated to the nucleolus and it dimerizes with the aryl hydrocarbon receptor nuclear translocator (ARNT), which together binds to the xenobiotic response element (XRE) site of

P4501A1 aryl hydrocarbon hydrolase (*CYP1A1*) located in the upstream enhancer region, which activates transcription. *CYP1A1* metabolizes toxic compounds converting them into hydrophilic intermediates which can be excreted. Therefore, *AHRR* hypomethylation with the consequent increased expression may reduce cellular responses to smoking through *CYP1A1*.[300] These results are in line with previous findings in which it seems that *AHRR* acts as a negative regulator of pathways involved in pleiotropic responses to environmental toxicants and its hypomethylation could be an adaptive response to this adverse exposure, controlling its detoxification.[150,302]

The MSDP-associated CpG with the lowest statistical significance and the most hypomethylation effect in child blood, cg25949550, which is located in *CNTNAP2* (*Contactin Associated Protein-Like 2*), was identified in prior studies to be hypomethylated in relation to current smoking[119,134] and MSDP,[140,303,304] having persistent dysregulation until adolescence.[140] CNTNAP2 is a cell adhesion protein whose decreased expression leads to aberrant neuronal migration and mutations in the gene have been associated with impaired language development, autism spectrum disorder and intellectual disability.[305] Since it is linked with neuronal problems, it is possible that MSDP and its persistent changes in methylation at cg25949550, might contribute to impaired neurodevelopment in exposed children.[306] *CNTNAP2* also contains another CpG (cg11207515), which was found to be hypermethylated in active smokers[306], whereas hypomethylated in cord blood[299], as well as in

our study in childhood blood. The different effect of current versus *in utero* exposure to smoking needs further investigation.

The second stronger hypomethylation effect was found in cg14179389, located in *GFI1* (*Growth Factor Independent 1 Transcriptional Repressor)*. *GFI1* contains another hypomethylated CpG that have been reported in most of the smoking studies during the prenatal period. The most hypermethylated CpG, cg12803068, is located in *MYO1G* (*Myosin 1G*) containing 3 more hypermethylated CpG in that loci.

The expression of *Teratocarcinoma-derived growth factor 1 (TDGF1*) was the only transcriptomic feature associated with. This gene encodes for an extracellular, membrane-bound signaling protein that plays an essential role in embryonic development and tumor growth.[307] It has found to be involved in cellular response to drug and regulation of heart rate by cardiac conduction pathway.[308] However, nothing is previously reported for smoking and further investigation is necessary to elucidate its role in MSDP effects. No more significant associations were found in gene expression in relation to MSDP despite having found associations between altered CpGs and expression of some genes through the eQTM analyses. Different reasons can explain the lack of association between MSDP and the child transcriptome: i) the effect of current smoking on gene expression is less persistent on time that the effect on DNA methylation,[300] ii) DNA methylation is not enough to change expression persistently, as many other mechanisms can also control

gene expression; iii) transcriptomics data might be more noisy than DNA methylation data because RNA is a less stable molecule than DNA. Of note, MSDP-associated child epigenetic marks could be important in triggering gene expression dysregulation in future postnatal exposures to tobacco smoking, both active and passive smoking. This requires further investigation in larger studies.

Finally, three urine metabolites (lactate, alanine and urea) were also found to be altered due to MSDP. However, after adjusting for postnatal SHS at home, which is related to recent exposures, they did not reach BN-significance and their effect was attenuated in 21% for lactate, 10% for alanine and 11% for urea. Therefore, these associations might be partially caused by postnatal exposure. Urine is the primary route by which the body eliminates water-soluble waste products and metabolites are downstream products of numerous genome-wide and protein-wide interactions and, generally, changes in urine composition are associated with acute exposures or body alterations. Elevated levels of lactate might be caused due to inhalation of carbon monoxide or cyanide and subsequent hypoxia.

5.3.2 Exposure to postnatal tobacco smoking and child molecular phenotypes

In contrast to MDSP, exposure to postnatal SHS, either assessed through questionnaire or cotinine, was mainly associated with child urinary and serum metabolites and plasma proteins, which are dynamic and changeable molecules, and thus might be capturing

more recent exposures. No associations were observed between SHS and the blood methylome once adjusted for MSDP, but we cannot discard that SHS has a small effect not detected with the actual sample size of this study.[137] Further investigation of the association of SHS on DNA methylation in larger sample sizes is needed.

We found lower concentrations of acyl-alkyl(ae)-phosphatidylcholines (PCs) in exposed children, which is consistent with findings in adults.[200] In agreement with our study, it has been reported that males that were current smokers had lower serum PC ae 38.0 and PC ae C40.6 levels.[200] Globally, comparing our results in children with finding in current adult smokers,[200] both age groups had lower levels of acyl-alkyl PCs if exposed to tobacco smoking, actively or passively. In general, we found decreasing levels of phosphatidylcholines and sphingomyelins, which are cell membrane molecules.

We also found that SHS increased plasma PAI1 (plasminogen activator inhibitor 1) protein levels, as it was previously found in current smoking.[309–311] PAI1 is encoded by the *SERPINE1* gene and it is the principal inhibitor of tissue plasminogen activator (tPA) and urokinase (uPA).[312] tPA and uPA are enzymes that convert plasminogen to plasmin to produce fibrinolysis, which is the fibrin conversion to soluble forms, and avoid thrombosis due to fibrin accumulation (Figure 41). High levels of PAI1 inhibit tPA and uPA, inhibiting fibrinolysis and increasing the risk of thrombosis. Thrombosis is an important risk factor for cardiovascular diseases

and can be one important cause of active cigarette smoking mediated by PAI1. Previous studies have reported higher levels of PAI1 in plasma of current smokers[309–311] as well as higher concentrations of plasminogen that results from inhibition of tPA and uPA by PAI1.[313] As far as we know this is the first time that PAI1 levels have been related to SHS. In active smokers, levels of PAI1 are approximately doubled but this novel finding reveals that SHS can also dysregulate plasma levels of this inhibitor increasing its levels nearly 30%. Therefore, although levels of exposure in passive smoking are lower than in active smoking, they are sufficient to induce changes at molecular level.

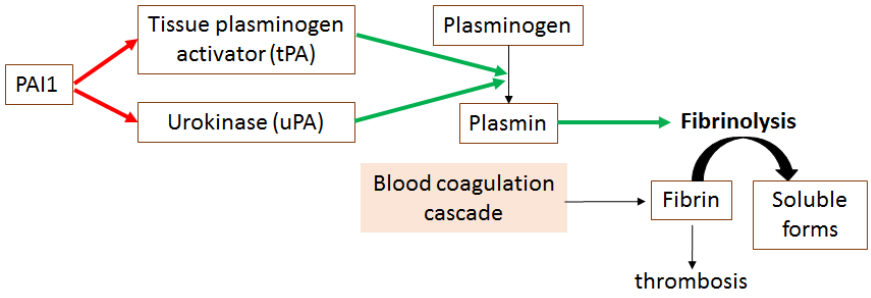

Figure 41. Schematic view of PAI1 cascade. In red inhibition steps and in green activation steps.

5.4 Dose, duration and frequency of the exposure

We investigated the effect of the dose, duration and frequency of the exposure to tobacco smoking. For prenatal smoking we used the following categories “no exposure”, “only mother-SHS exposure”,

“non-sustained smoker”, “sustained smoker at low dose (=<9)” and “sustained smoker at high dose (>9)”; and for postnatal exposure: “no exposure”, “exposure to SHS only outside home”, “exposure to SHS only inside home” and “exposure to SHS inside and outside home”. Given that children expend a higher number of hours at home than in the same place outside home, we assume higher exposure levels for those children which are exposed at home, than for those that are exposure outside home. Indeed, cotinine levels, a measure of recent exposure, showed higher correlation with home-SHS than with outside-SHS.

Our studies on placental DNA methylation found that increased duration of exposure (e.g. sustained- versus any-MSDP) produced larger magnitudes of associations, however in placenta not all the participants had enough data to contribute to the sustained-MSDP models and we were unable to evaluate whether identified relationships demonstrated dose-response patterns. Thus, additional follow-up studies should examine dose-response relationships at placental identified loci, using reported cigarettes per day as we did in the study of blood methylation, and investigate if there are critical windows during pregnancy in which the placental epigenome is more susceptible to MSDP.

In child blood, we also showed that differential DNA methylation was greater with increased duration of the exposure, finding larger effects in children exposed to sustained-MSDP in comparison to those exposed to any-MSDP, as occurred in placenta tissue at birth.

Moreover, considering SHS of the mother and the number of cigarettes smoked during pregnancy we showed a dose-effect response for some CpGs, although in others we only saw dose-effect for sustained-MSDP with increasing number of cigarettes, or a saturated pattern at certain point, whereas for some other we could not identify a clear pattern. This suggests that not all the CpGs respond in the same way when increasing dose or duration of smoking, but it also could be that some CpGs were measured with less precision.

For postnatal exposure to SHS, considering if it occurs at home, outside or both places, we showed a dose-effect response for PAI1 protein levels, whereas we did not identify clear patterns for other metabolites.

Given that some of the molecular features analyzed in the study show a dose-response effect, they could be used as novel biomarkers of exposure to tobacco smoking. In addition, those in DNA methylation loci could be used to track prenatal exposures as the dose-response effects were persistent, at least until childhood. Initial analysis using DNA methylation to develop a biomarker of MSDP have been published.[26]

5.5 Tissue specific responses to exposure to tobacco smoking: cord blood and placenta

We compared our findings in placenta to those of a previous PACE meta-analysis of MSDP and cord blood DNA methylation.[304] Six CpG sites [cg05575921 and cg21161138 (*AHRR*), cg25189904 (*GNG12*), cg15893360 (*PXN*), cg16547579 (*SLC23A2*), and cg07565956 (*ZBTB4*)], that were identified in the previous PACE analysis were reproducible at the BN-significant level and with the same direction in our placental DNA methylation meta-analysis. Thus, the epigenomic response of *AHRR*, *GNG12*, *PXN*, *SLC23A2*, and *ZBTB4* to prenatal tobacco smoke exposure may be somewhat consistent across fetal tissues. This is the case also for cg05575921 (*AHRR*), despite of being located in a maternal placenta-specific gDMR.[222] However, its MSDP associated methylation change in placenta was ½ of that observed in cord blood, while in contrast, cg21161138, also at *AHRR* but outside the gDMR, showed similar methylation change in both tissues. However, the minimal overlap in the overall MSDP-associated loci between the placenta and cord blood likely represent unique tissue-specific biological responses to this toxic exposure. We also detected opposite relationships between MSDP and DNA methylation in placenta and cord blood in some genes. For instance, MSDP associates with decreased DNA methylation at cg08327744 in *RNF122* in cord blood, and higher DNA methylation in the placenta. While MSDP associates with higher DNA methylation at cg12101586 and cg23680900 in *CYP1A1*, but with lower DNA methylation in placenta. Higher

expression of *CYP1A1* has been found in placentas of smoker mothers.[175] The opposite direction of the effects in these two tissues in these genes requires further investigation, but we anticipate that placenta is an extraembryonic tissue with a different developmental origin and methylation reprogramming.[112,222,314]

Given the difficulty of accessing to certain tissues, there is not enough data to have a comprehensive understanding of the tissue specific responses to environmental pollutants. As far as we know, only few studies have compared the DNA methylation signature of tobacco smoking among different tissues. One of them compared blood, skin, lung and adipose tissue.[300] Although the authors indicate that there are smoking associated loci in adipose tissue which are mirrored in blood, skin, and lung; it is also true that there were others that were tissue-specific. For instance, in our study, 9 CpGs located in *CYP1A1* were hypermethylated in child blood in response to MSDP and also in cord blood.[299] Oppositely, cg12101586 (*CYP1A1*) was found to be hypomethylated in placenta as we have shown in our results, and previously reported.[152] Interestingly, an increase in *CYP1A1* mRNA expression in placenta was observed during the first trimester of pregnancy, which may indicates that the placenta mediates an intermediate defense response through *CYP1A1* activation to detoxify the xenochemicals from MSDP. In adipose tissue, current tobacco smoking is related to *CYP1A1* hypomethylation, in line with the pattern in placenta. These findings show that smoking produces different responses in different tissues.[315]

If future studies confirm tissue-specific responses to environmental pollutants, this will have some implications in the field of epigenetic environmental epidemiology. The main contributions of epigenetics in environmental epidemiology are: i) the understanding of the molecular mechanisms that link the exposure to the disease outcome; and ii) the development of biomarkers of exposure. While DNA methylation in blood, the most widely studied tissue, will still be useful for the development of biomarkers of exposure, some concerns will appear about its use in the understanding of the molecular mechanisms underlying the effect of environmental exposures on disease outcomes in tissues other than the immune system. These biomarkers could help, for example, in childhood health studies filling missing data from self-reported questionnaires or validating it.[26]

5.6 Tissue cellular composition

Placenta is a heterogeneous tissue with multiple different cell types[316] that serve different functions in the development and activities of the placenta, and thus have different epigenetic states.[317] To correct for this, we estimated and adjusted for variability in placental DNA methylation that may be due to tissue heterogeneity. We utilized a data driven approach that was not based on a methylome reference, as no references for placental cell-type methylomes are currently available, nor measured heterogeneity via cell sorting, as this was not

feasible for this study. Adjustments for these estimates of putative cellular heterogeneity did reduce heterogeneity in the meta-analyses, thus improving the consistency of observed associations between MSDP and DNA methylation across these independent cohorts. However, our adjustments for putative tissue heterogeneity did not improve genomic inflation as estimated by lambda, and it is possible that residual confounding may have influenced some of our results.

In blood, which is also an heterogenous tissue containing different cell types, we could estimate cell proportions using the Houseman algorithm and the Reinius reference panel from DNA methylation data.[268,269] For RNA data (gene and miRNA expression) we used a reference free approach, the surrogate variable analysis[256], which assumes that maximum variability of the data is due to cell heterogeneity and technical batch effects.

In cellular heterogeneous tissues it is hard to disentangle if smoking induced methylation changes are cell-type specific or if the change is quite homogeneous through all cell types. Bauer, *et al.* carried of some studies[100,318] demonstrating that exposure to tobacco smoking can induce expansion of specific immune cell types as a consequence of, for example, lung inflammation. For instance, hypomethylation at cg05575921 (*AHRR*), seemed to be found in different cell types, with the most pronounced methylation change in granulocytes. In contrast cg19859270 at GPR15 was mainly explained by an expansion of a particular CD3+ T memory cell.

The non-homogeneous effect of exposures on DNA methylation across blood cell types could explain the small effect sizes observed in EWAS.[114] The reason why some smoking associated DNA methylation signatures remaining persistent over time while other are erased faster, is not clear, but a possible explanation could be cellular heterogeneity. Blood cell type proportions change physiologically from birth to childhood, with a decrease in granulocytes and an increase in lymphocytes; and, as commented above, some of the smoking effects seem to be specific or more predominant in particular immune cell types, such as GPR15+ effector memory T cells, which have a longer life span and we would expect that methylation of CpGs in this cell type could be interpreted as persistent. However, the interpretation of persistent CpGs mainly found in granulocytes remains unclear (e.g. cg09935388 (*GFI1*)).

5.7 Strengths and limitations

The findings from this thesis should be considered within the context of the studies' limitations. First, exposure assessment to tobacco smoking through either questionnaire or cotinine have some intrinsic limitations. Maternal smoking definitions and child exposure to SHS were self-reported or assessed through parental-reported questionnaire data, and they are subjected to misreporting. Child exposure to SHS, was also assessed measuring cotinine, the main metabolite of nicotine, which is easily detected in urine. However, its half-life is of approximately 20h, giving information only about the

most recent exposure and it does not allow a clear discrimination of prolonged SHS exposed individuals.[190,192,193]

Second, our studies predominantly consisted of samples from mother-infant pairs of European ancestry, and thus additional studies involving diverse ethnic backgrounds are needed in order to improve the generalizability of these findings.

Third, we cannot infer fully on causality and directions of effect; while it is unlikely that DNA methylation, or other biomarkers, would influence maternal smoking, our associations between placental methylation levels and reproductive outcomes could be due to reverse causation. Two-sample Mendelian randomization analysis could be used to investigate the causal link between placental DNA methylation and birth outcomes.[319] On the other hand, although the statistical models were adjusted for a list of a priori confounders (in the placenta study) or confounders selected though DAGs (in the HELIX study), we cannot completely rule out residual confounding. Residual confounding could be addressed by using paternal smoking during pregnancy as a negative control, only if no passive smoking can be proved.

Despite these limitations, our studies had numerous strengths. In the placental DNA methylation study, we had a large overall sample-size and seven independent studies to identify these relationships. We used harmonized definitions of exposure variables and covariates, standardized protocols for quality control and pre-processing of DNA

methylation data, and standardized methods for statistical analyses. We also performed secondary analyses involving gene expression, functional and phenotype enrichment, overlap with GWAS hits for reproductive outcomes, and meta-analyses of DNA methylation variation with birth outcomes to provide comprehensive biological and health-related interpretations of these findings.

In the HELIX study, information on prenatal and postnatal (8 years) exposure to tobacco smoking was available for a large sample size. Moreover, this is a unique study including data from five different molecular layers (methylation, transcription, miRNA expression, proteins and metabolites in two biological matrices) for which samples for all the cohorts were processed in the same laboratory and though a randomization step.

Given the adverse effects of tobacco smoking in health and the possible mediation through molecular mechanisms, this study provides a global vision of different molecular marks and how they are altered in children due to either MSDP or SHS being able to detect short and long-term effects. Our results suggest that MSDP altered blood DNA methylation patterns in a persistent way, whereas the proteome and metabolome were indicators of more recent exposure to SHS.

6 CONCLUSIONS

1. In our studies, approximately 20% and 10% of the children were exposed to any and sustained maternal smoking during pregnancy (MSDP), respectively. Exposure raised to 35% for postnatal secondhand smoking (SHS). The correlation between prenatal and postnatal exposure ranged from 0.4 to 0.7, and marked differences were observed among countries.

2. In placenta, exposure to MSDP was associated with differentially methylated patterns. Significant CpGs were enriched for enhancers and for pathways related to inflammation, growth factors and hormone signaling, and vascularization and angiogenesis.

3. Many of the CpG sites associated to MSDP in placenta were also related to placental gene expression, gestational duration and birth size, suggesting a potential mediating role. For instance, methylation of cg27402634, upstream *LEKR1*, was dramatically hypomethylated in response to MDSP, and associated with *LEKR1* gene expression and lower birth size.

4. Exposure to MSDP was associated with differential blood DNA methylation patterns at age 6-12 years, indicating the persistence of epigenetic marks; whereas child exposure to SHS was associated with cross-sectional changes in protein and metabolite levels, suggesting that the dynamic nature of these molecules might reflect more recent exposures.

5. We confirmed previously reported associations between MSDP and child blood DNA methylation in 35 CpGs. In addition, we identified 15 expression quantitative trait methylation (eQTM) loci for these CpGs, although no direct effect of MSDP on transcription was observed, suggesting that changes in gene expression are less persistent.

6. Child exposure to SHS was related to increased levels of PAI1, which promotes thrombosis, and to altered metabolic profiles. Similar findings have been reported in current adult smokers. Therefore, although levels of exposure in passive smoking are lower than in active smoking, they are sufficient to induce changes at molecular level.

7. In both tissues, sustained-MDSP had a stronger effect on DNA methylation than any-MSDP. Moreover, levels of some child molecular features showed an increased or decreased tendency with increasing dose and duration of exposure to MSDP (some blood CpGs) or to SHS (PAI1).

8. Minimal overlap has been found between the response to MSDP in placenta and cord blood tissues, suggesting unique tissue-specific biological responses to toxic exposures.

SUPPLEMENTARY MATERIAL

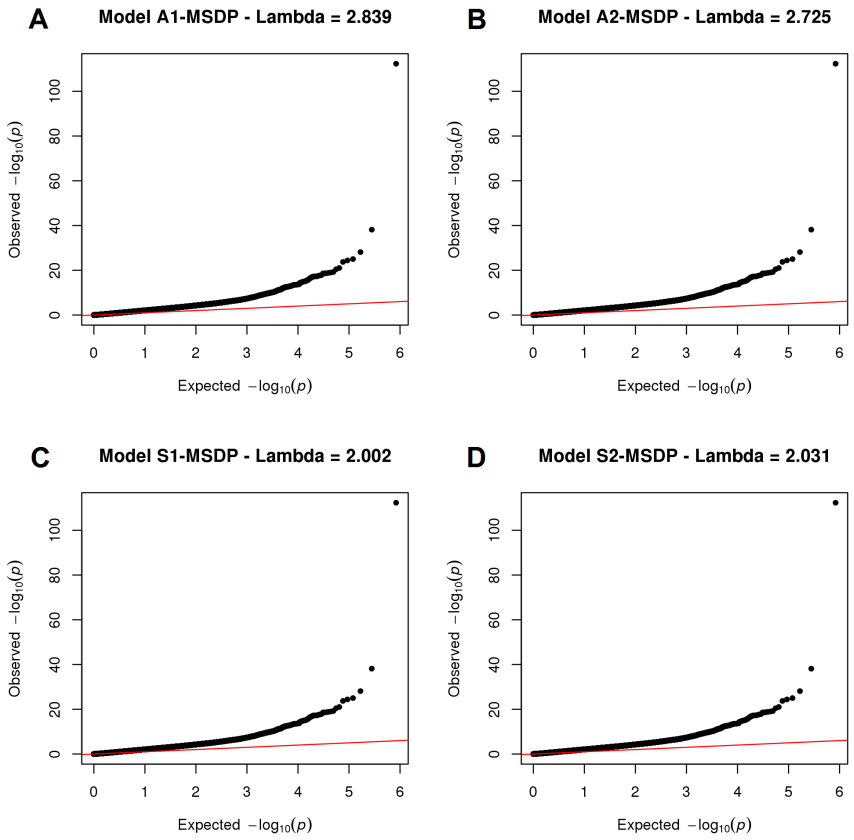

Figure S1. QQ-plot for the meta-analyses of the association between placental DNA methylation and any MSDP without (A) and with adjustment for cellular heterogeneity (B) and sustained MSDP without (C) and with adjustment for cellular heterogeneity (D). N=1700 (A and B) and N=795 (C and D).

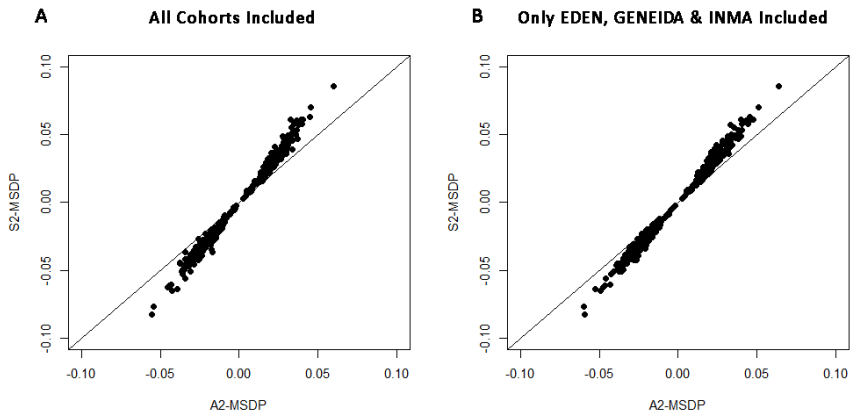

Figure S2. Plots showing the increase in magnitude of coefficients for model S2 versus A2, (A) when A2 pooled estimates included all particiapting cohorts, and (B) when A2 pooled estimates only included those cohorts that also participated in model S2 (EDEN, GENEIDA and INMA); this analysis included the 547 CpGs that were Bonferroni-sgnificant in models A2 and S2. One CpG (cg27402634) excluded from plot due to being an outlier and making visualization of the overall differences between A2 and S2 more difficult; though this site also showed a larger magnitude of effect in model when all cohorts were included (difference between S2 and A2 = 12.7%), and when only EDEN, GENEIDA, and INMA were included (difference between S2 and A2 = 27.5%).

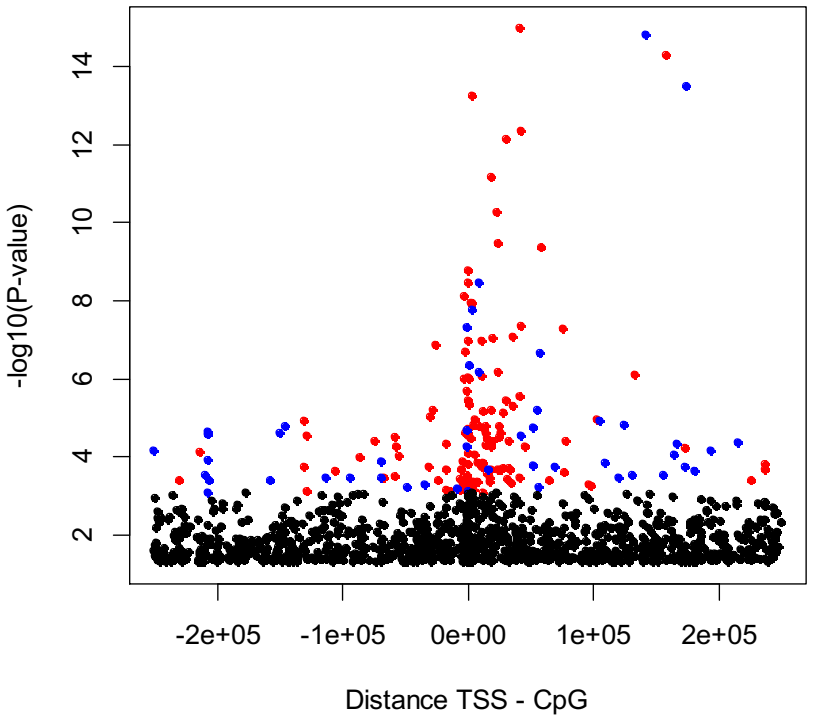

Figure S3. –log10(P-value) versus distance between gene transcription start site (TSS) and CpG site for 1219 eQTMs. In blue positively associated eQTMs at 5% FDR, and in red inversely associated eQTMs at 5% FDR.

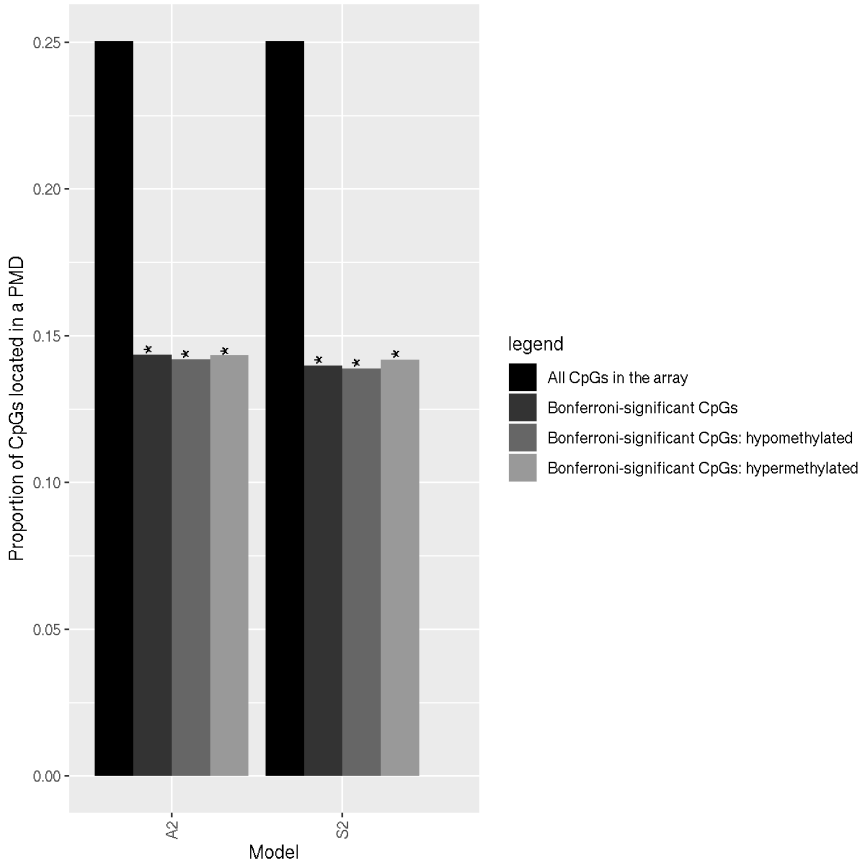

Figure S4. Proportion of CpGs annotated to placental partially methylated domains (PMDs) in the A2-MSDP model and in the S2-MSDP model.

*Nominally significant P-value in the hypergeometric test

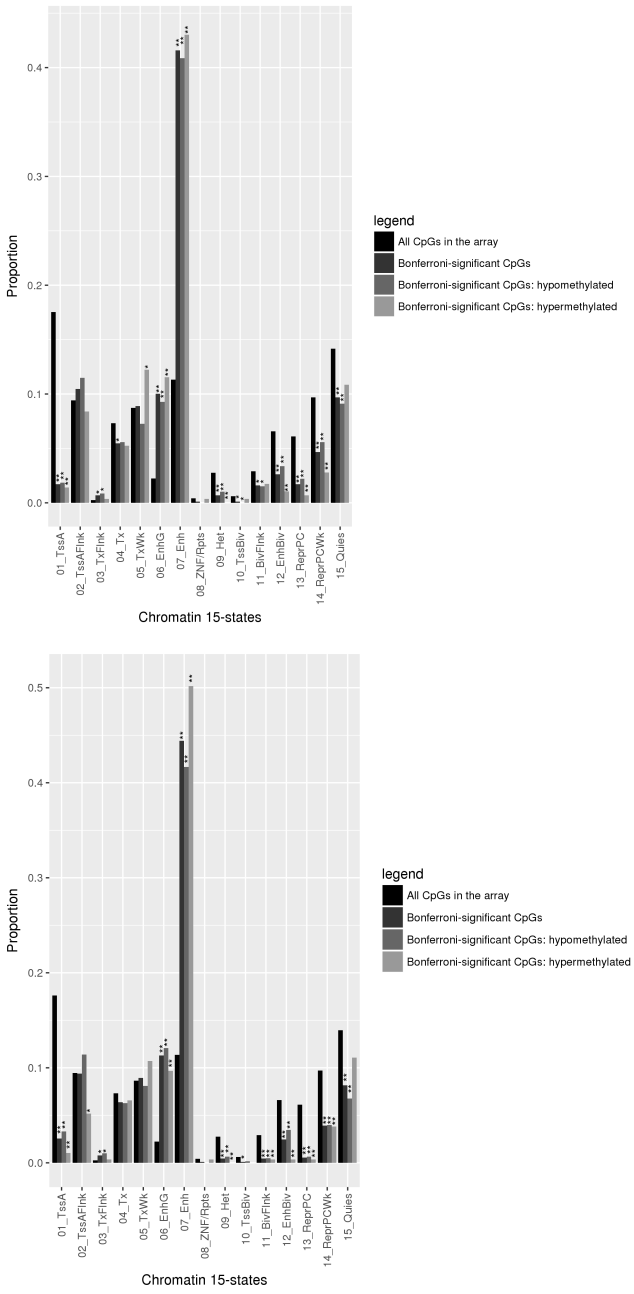

Figure S5. Proportion of CpGs annotated to each of the 15 placenta chromatin states from ROADMAP, by significance and direction of the effect in the A2-MSDP model (A) and in the S2-MSDP model (B).

Chromatin states are defined as: 01_TssA = Active transcription start site; 02_TssAFlnk = Flanking active transcription start site; 03_TxFlnk = transcribed state at the 5' and 3' end of genes showing both promoter and enhancer signatures; 04_Tx = Actively transcribed state (strong transcription); 05_TxWk = Actively transcribed state (weak transcription); 06_EnhG = Genic enhancer; 07_Enh = Enhancer; 08_ZNF/Rpts = state associated with zinc finger protein genes and repeats; 09_Het = Consitutive heterochromatin; 10_TssBiv = Bivalent/Poised TSS; 11_BivFlnk = Flanking bivalent TSS/Enh; 12_EnhBiv = Bivalent enhancer; 13_ReprPC = repressed polycomb; 14_ReprPCWk = Weak repressed polycomb; 15_Quies = Quiescent state.

*Nominally significant P-value in the hypergeometric test

**Bonferroni significant P-value (p-value< 0.003)

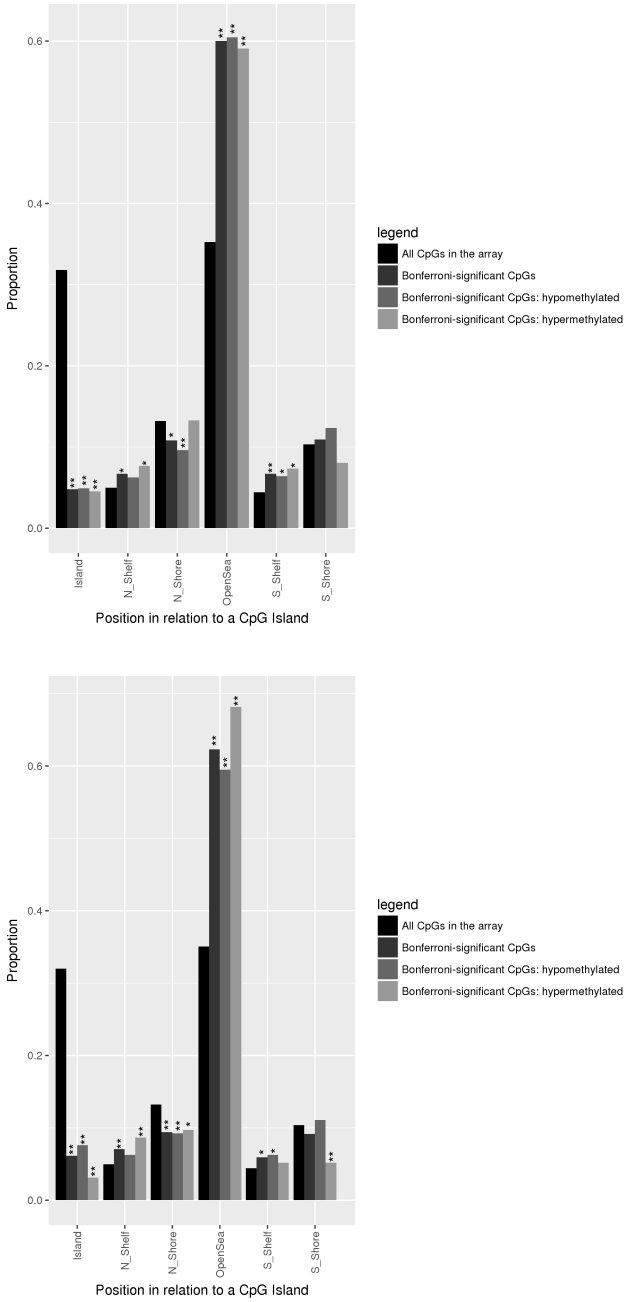

Figure S6. Proportion of CpGs annotated to each of the 6 locations in relation to a CpG island, by significance and direction of the effect in the A2-MSDP model (A) and the S2-MSDP model (B).

*Nominally significant P-value in the hypergeometric test

**Bonferroni significant P-value (p-value<0.008)

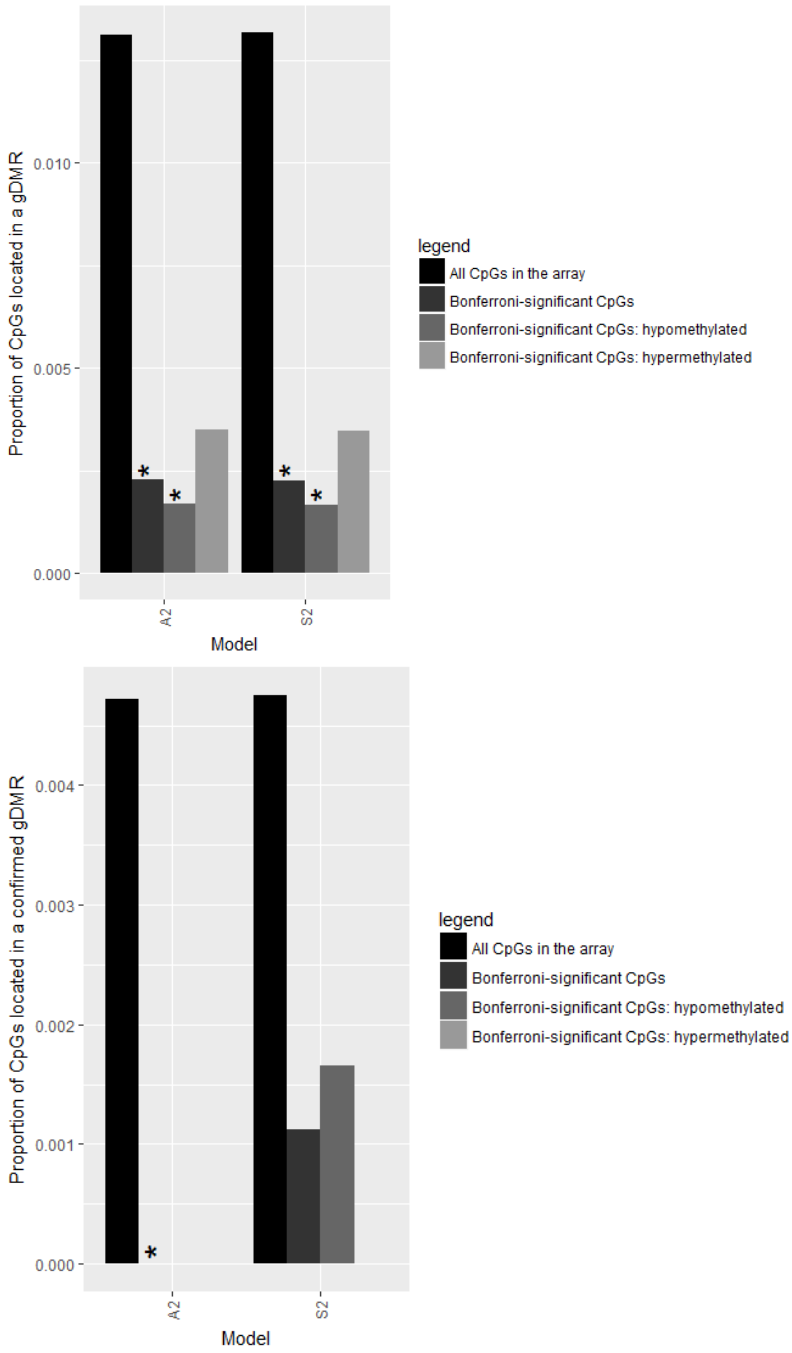

Figure S7. Proportion of CpGs annotated to candidate (A) or confirmed (B) placental germline differentially methylated regions (gDMR) in the A2-MSDP model and in the S2-MSDP model.

*Nominally significant P-value in the hypergeometric test

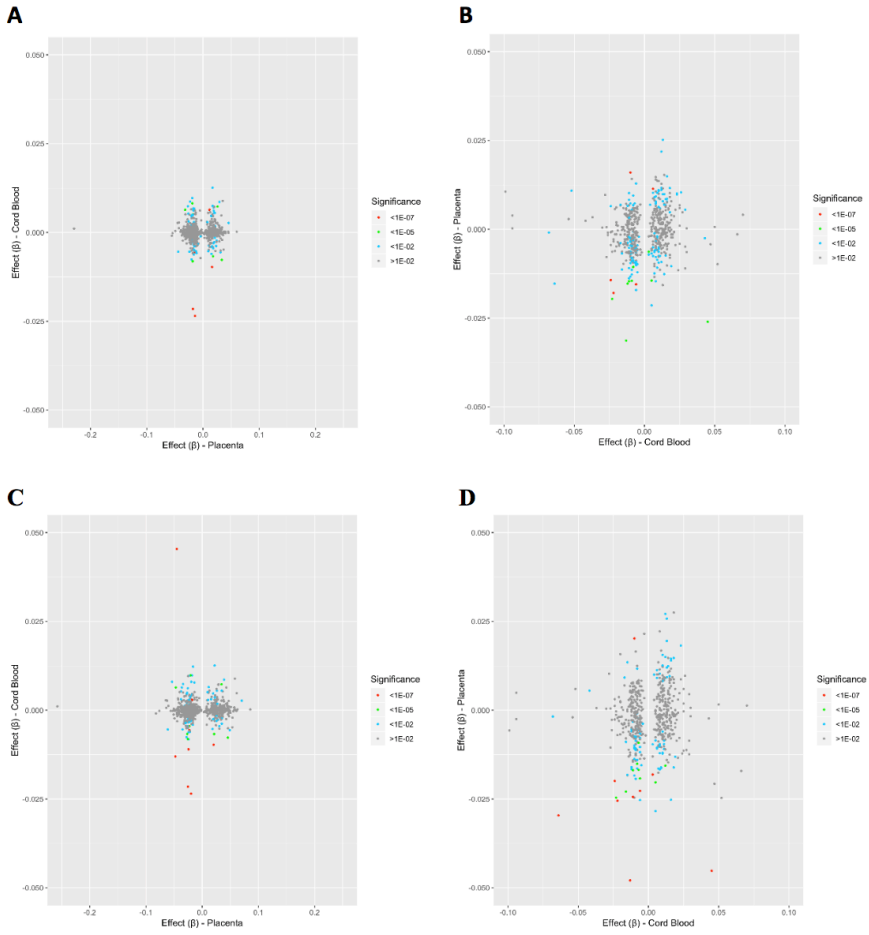

Figure S8. Correlation of regression coefficients of MSDP on DNA methylation in placenta vs. cord blood. The x-axis shows the coefficients of 5% FDR-significant CpGs in a particular tissue, while the y-axis shows the equivalent coefficients in the other tissue. Statistical significance of the regression coefficients of the tissue in the y-axis are indicated with different colors, being BN-significant CpGs shown in red (P-value<1E-07). Cord blood coefficients are for sustained MSDP and not adjusted for blood cell type proportions. (A) Coefficients of FDR-significant CpGs in placenta (model A2) vs coefficients in cord blood. Pearson correlation R2=0.04 (R2=0.05 without top CpG site). (B) Coefficients of FDR-significant CpGs in cord blood vs coefficients in placenta (model A2). Pearson correlation R2=0.09. (C) Coefficients of FDR-significant CpGs in placenta (model S2) vs coefficients in cord blood. Pearson correlation R2=0.04 (R2=0.05 without top CpG site). (D) Coefficients FDR-significant CpGs in cord blood vs coefficients in placenta (model S2). Pearson correlation R2=0.08. Similar results of no correlation were obtained with different

smoking definitions for cord blood and different cellular heterogeneity adjustment in placenta (data not shown).

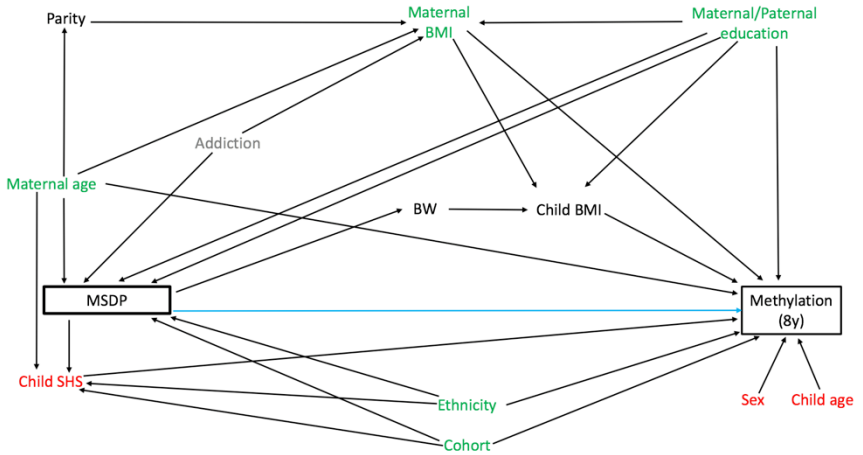

Figure S9. DAG diagram with causal assumptions from a priori knowledge between MSDP and methylation. In green, covariates proposed as confounders by DAGitty tool.[264] In red, covariates not proposed by DAGitty but included in all or some models. In black, covariates not proposed by DAGitty tool and not included in the models. In grey, not measured covariates. In blue, the causal path assessed in the model.

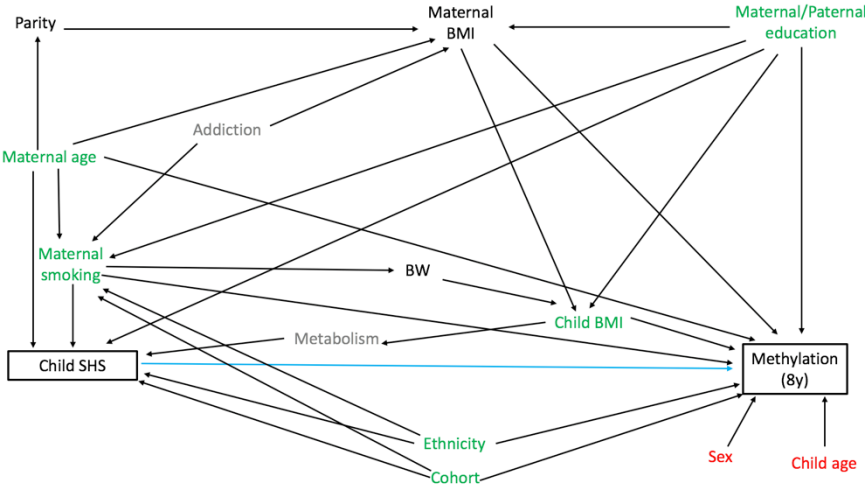

Figure S10. DAG diagram with causal assumptions from a priori knowledge between child SHS and methylation. In green, covariates proposed as confounders by DAGitty tool.[264] In red, covariates not proposed by DAGitty but included in all or some models. In

black, covariates not proposed by DAGitty tool and not included in the models. In grey, not measured covariates. In blue, the causal path assessed in the model.

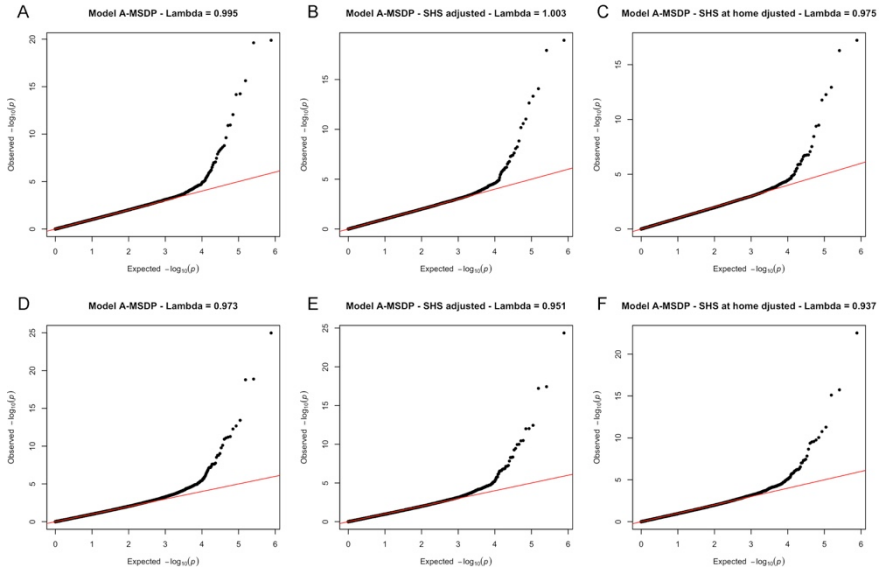

Figure S11. QQ-plot for the associations between DNA methylation at 8 years and A-MSDP without (A) and with postnatal SHS adjustment (B) and with postnatal SHS at home adjustment (C); and the associations between DNA methylation at 8 years and S-MSDP without (D) and with postnatal SHS adjustment (E) and with postnatal SHS at home adjustment (F).

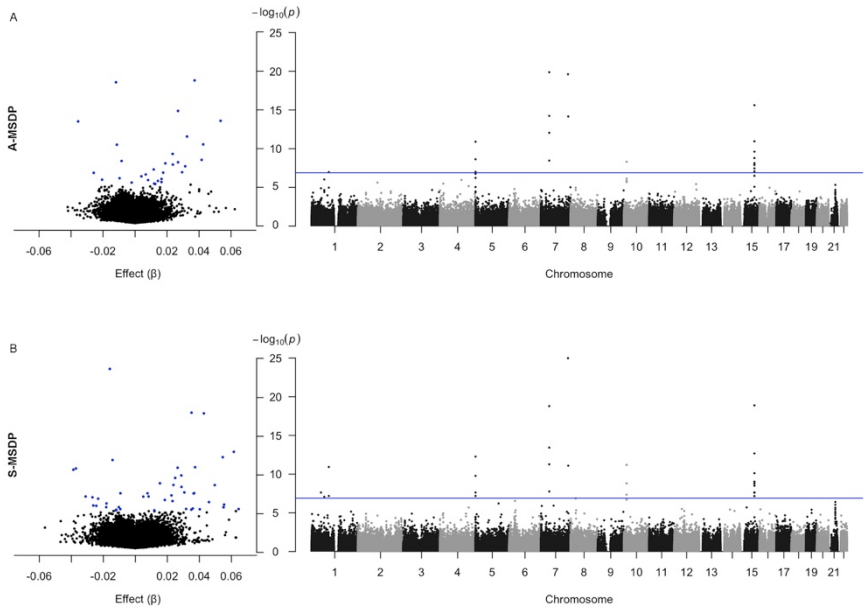

Figure S12. Volcano and Manhattan plots of the association between A-MSDP (A) and S-MSDP (B) and DNA methylation at 8 years. In the volcano plots, the x-axis shows the change in the DNA methylation (effect) and the y-axis the statistical significance (-log10 (P. Value)). 5% FDR-significant CpGs are shown in blue. Among them, 26.7% (A) and 34.8% (B) were hypomethylated. Sample sizes: N=1105 (A) and N=1047 (B). The Manhattan shows the -log10 (P. Value) in the y-axis and the genomic location of the CpGs in the chromosomes in the x-axis. The blue line represents the BN-significance threshold.

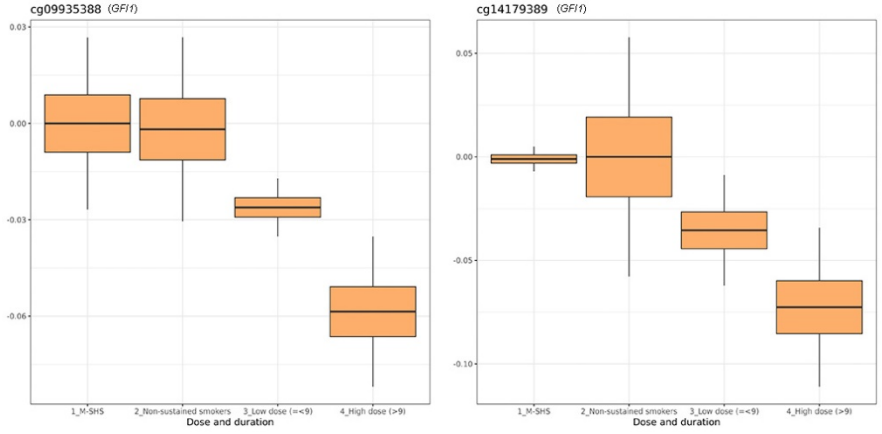

Figure S13. Box Plots showing the change of methylation in relation to control group (non-exposed mothers) in CpGs located near *GFI1* gene. Sample size: 0_non-exposed (N=582);

1_M-SHS (N=351); 2_Non-sustained smokers (N=58); 3_Low dose (=<9 cig/d) (N=89); High dose (>9 cig/d) (N=15).

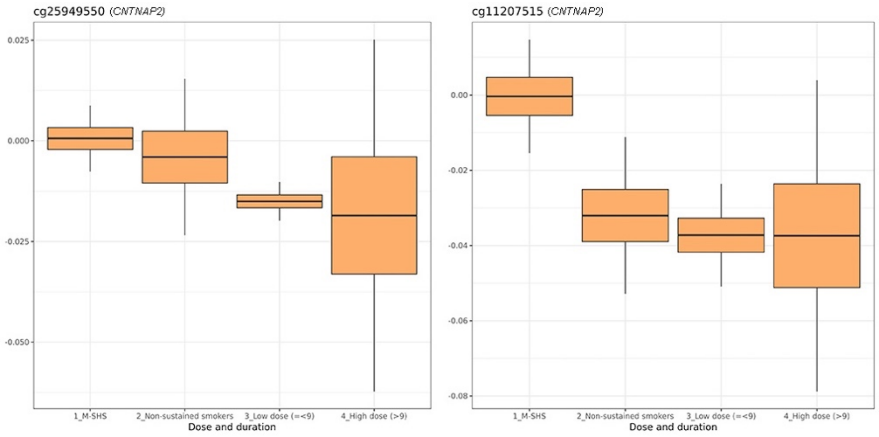

Figure S14. Box Plots showing the change of methylation in relation to control group (non-exposed mothers) in CpGs located near *CNTNAP2* gene. Sample size: 0_non-exposed (N=582); 1_M-SHS (N=351); 2_Non-sustained smokers (N=58); 3_Low dose (=<9 cig/d) (N=89); High dose (>9 cig/d) (N=15).

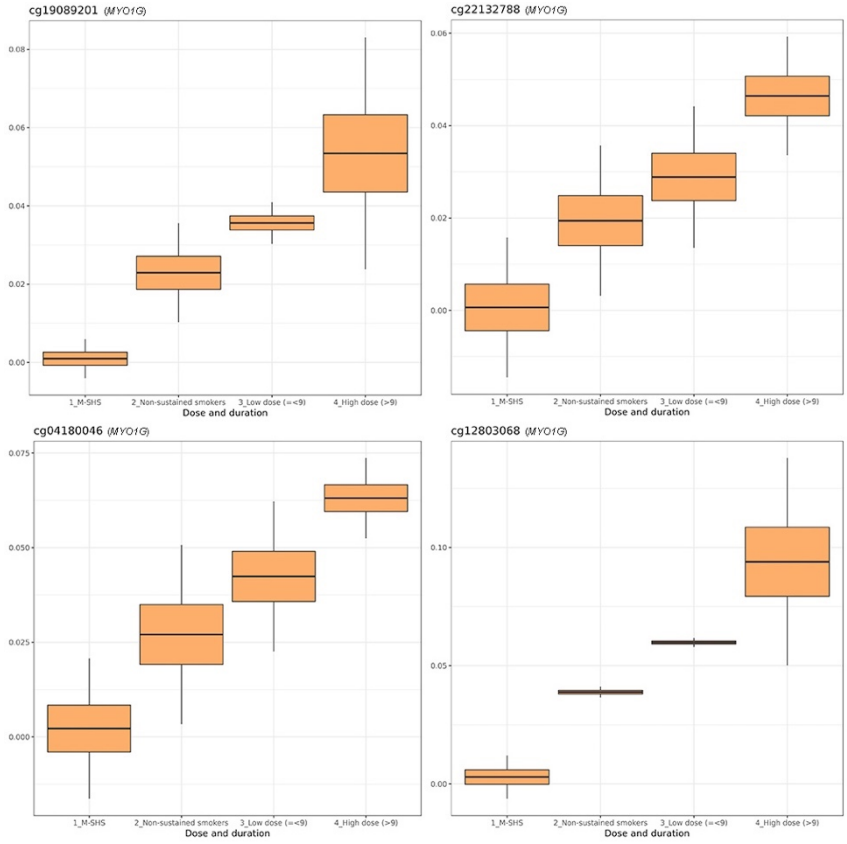

Figure S15. Boxplots showing the change of methylation in relation to control group (non-exposed mothers) in CpGs located near MYO1G gene. Sample size: 0_non-exposed (N=582); 1_M-SHS (N=351); 2_Non-sustained smokers (N=58); 3_Low dose (=<9 cig/d) (N=89); High dose (>9 cig/d) (N=15).

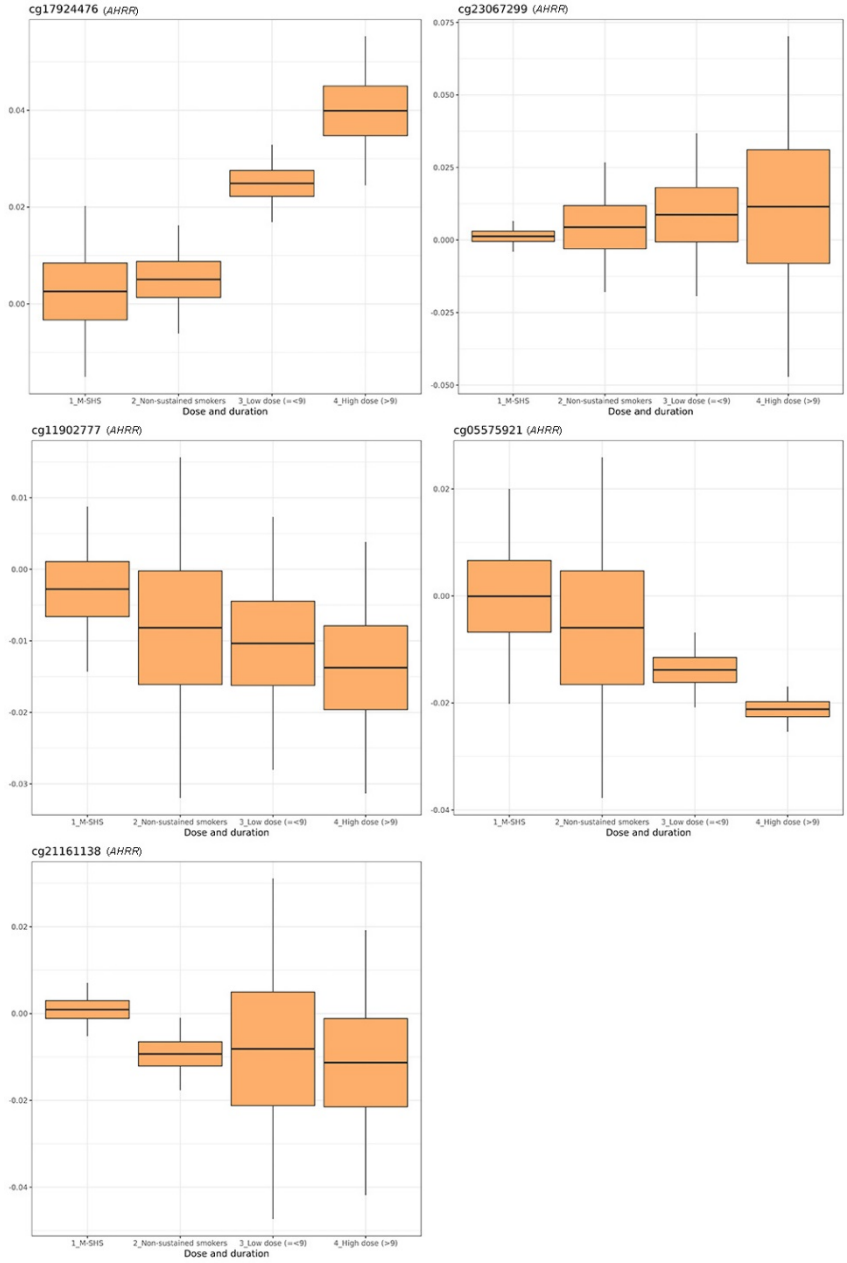

Figure S16. Boxplots showing the change of methylation in relation to control group (non-exposed mothers) in CpGs located near AHRR gene. Sample size: 0*_non-exposed* (N=582); 1_M-SHS (N=351); 2_Non-sustained smokers (N=58); 3_Low dose (=<9 cig/d) (N=89); High dose (>9 cig/d) (N=15).

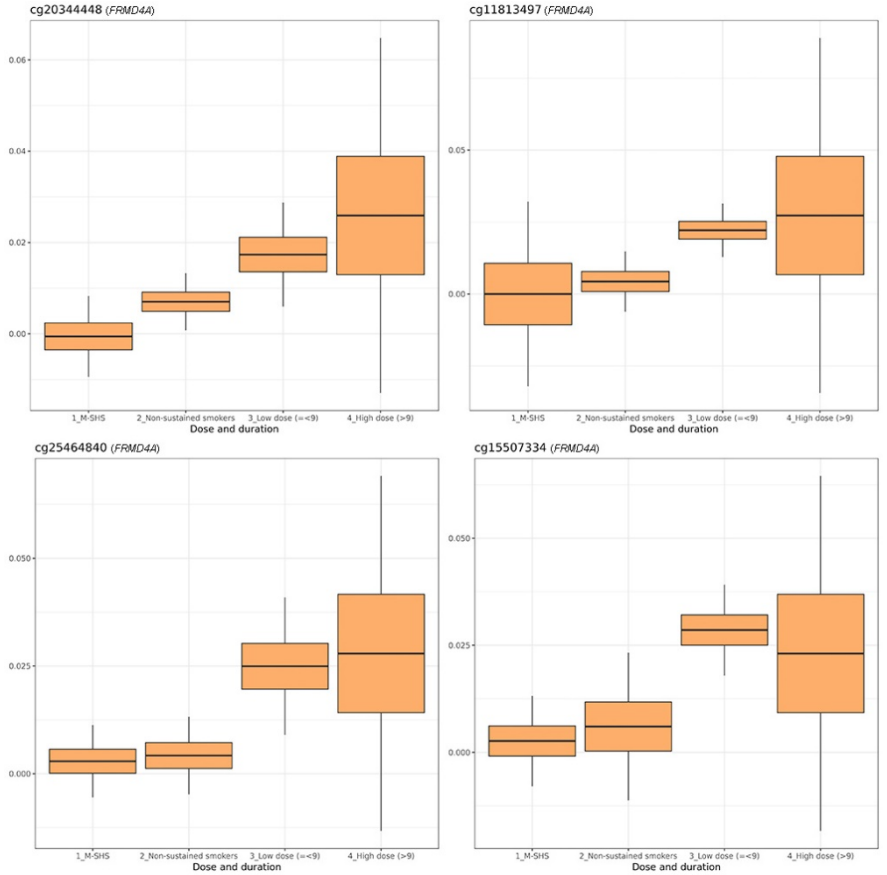

Figure S17. Boxplots showing the change of methylation in relation to control group (non-exposed mothers) in CpGs located near *FRMD4A* gene. Sample size: 0_non exposed (N=582); 1_M-SHS (N=351); 2_Non-sustained smokers (N=58); 3_Low dose (=<9 cig/d) (N=89); High dose (>9 cig/d) (N=15).

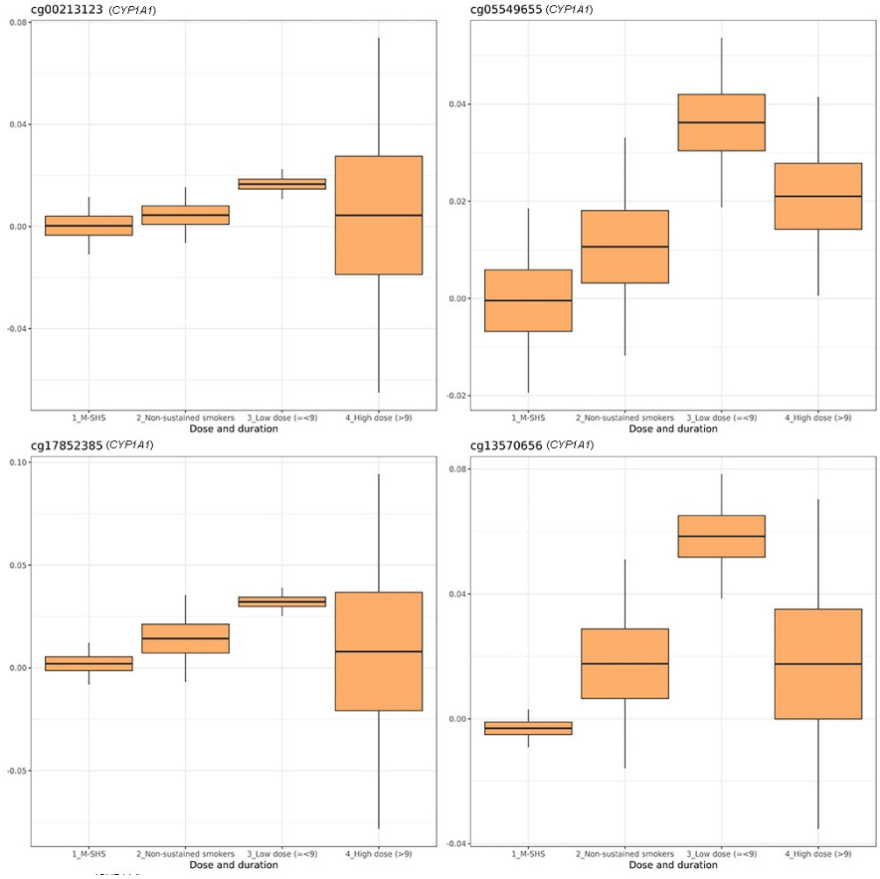
cg00213123 (CYP1A1)
0.08
0.04
0.00
-0.04
1_M-SHS
2_Non-sustained smokers
3_Low dose (=<9)
4_High dose (>9)
Dose and duration
cg05549655 (CYP1A1)
0.04
0.02
0.00
-0.02
1_M-SHS
2_Non-sustained smokers
3_Low dose (=<9)
4_High dose (>9)
Dose and duration
cg17852385 (CYP1A1)
0.10
0.05
0.00
-0.05
1_M-SHS
2_Non-sustained smokers
3_Low dose (=<9)
4_High dose (>9)
Dose and duration
cg13570656 (CYP1A1)
0.08
0.04
0.00
-0.04
1_M-SHS
2_Non-sustained smokers
3_Low dose (=<9)
4_High dose (>9)
Dose and duration

Figure S18. Boxplots showing the change of methylation in relation to control group (non-exposed mothers) in CpGs located near *CYP1A1* gene. Sample size: 0_non-exposed (N=582); 1_M-SHS (N=351); 2_Non-sustained smokers (N=58); 3_Low dose (=<9 cig/d) (N=89); High dose (>9 cig/d) (N=15).

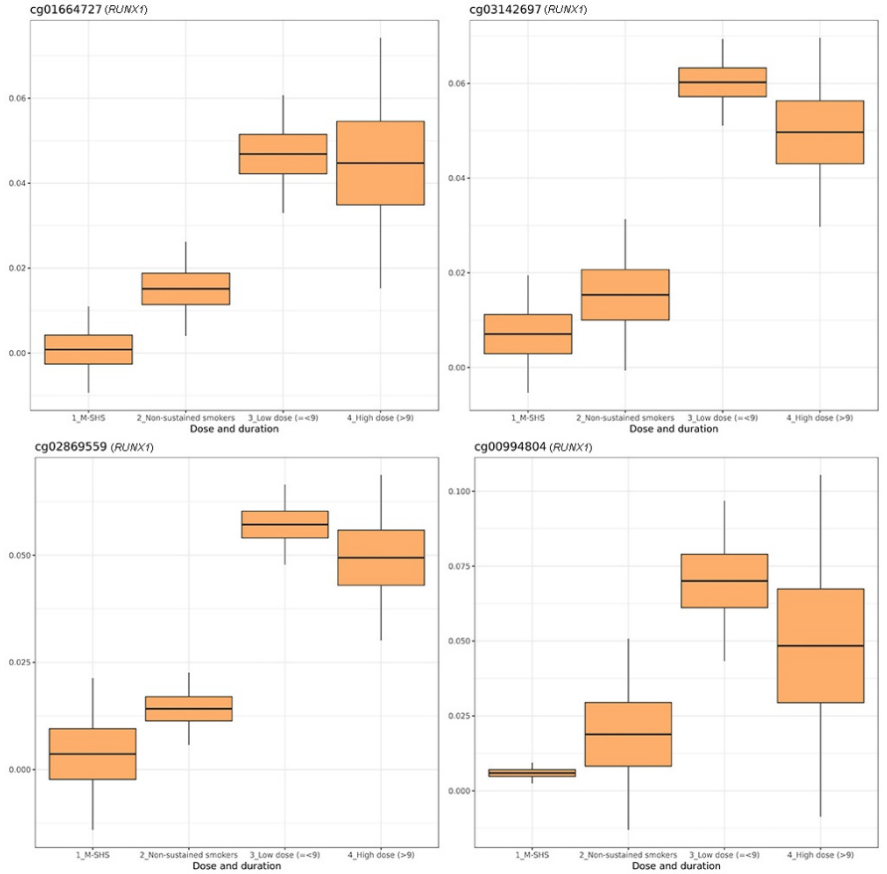

Figure S19. Boxplots showing the change of methylation in relation to control group (non-exposed mothers) in CpGs located near *RUNX1* gene. Sample size: 0_non-exposed (N=582); 1_M-SHS (N=351); 2_Non-sustained smokers (N=58); 3_Low dose (=<9 cig/d) (N=89); High dose (>9 cig/d) (N=15).

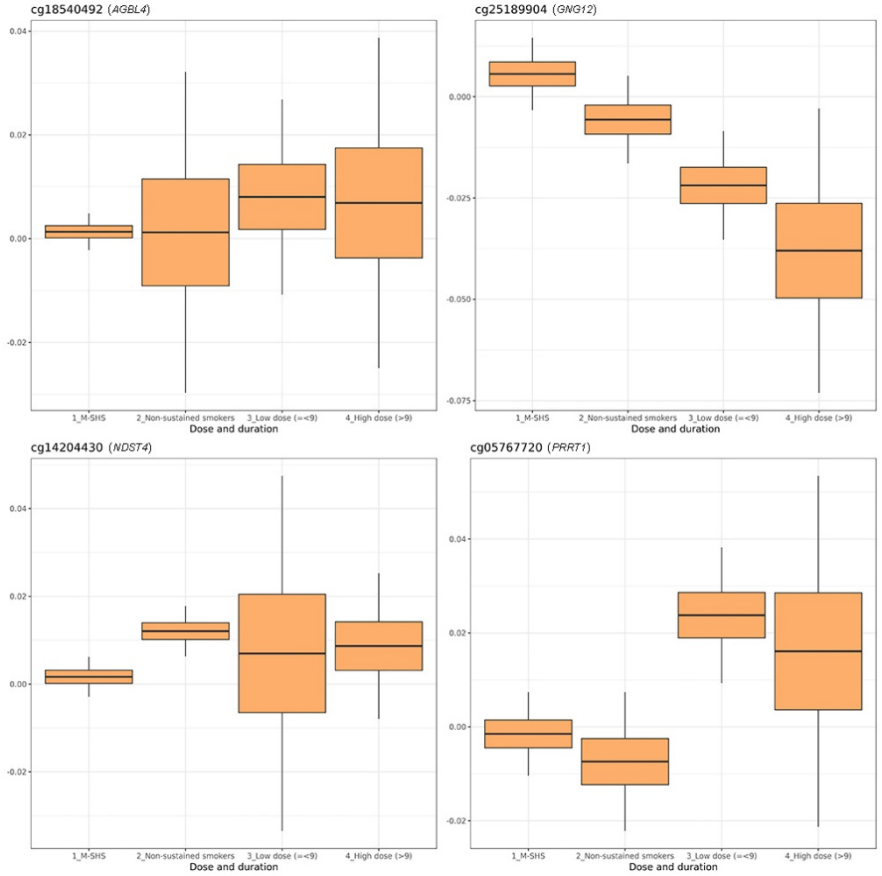

Figure S20. Boxplots showing the change of methylation in relation to control group (non-exposed mothers) in CpGs located near *AGBL4, GNG12, NDST4* and *PRRT1*. Sample size: 0_non-exposed (N=582); 1_M-SHS (N=351); 2_Non-sustained smokers (N=58); 3_Low dose (=<9 cig/d) (N=89); High dose (>9 cig/d) (N=15).

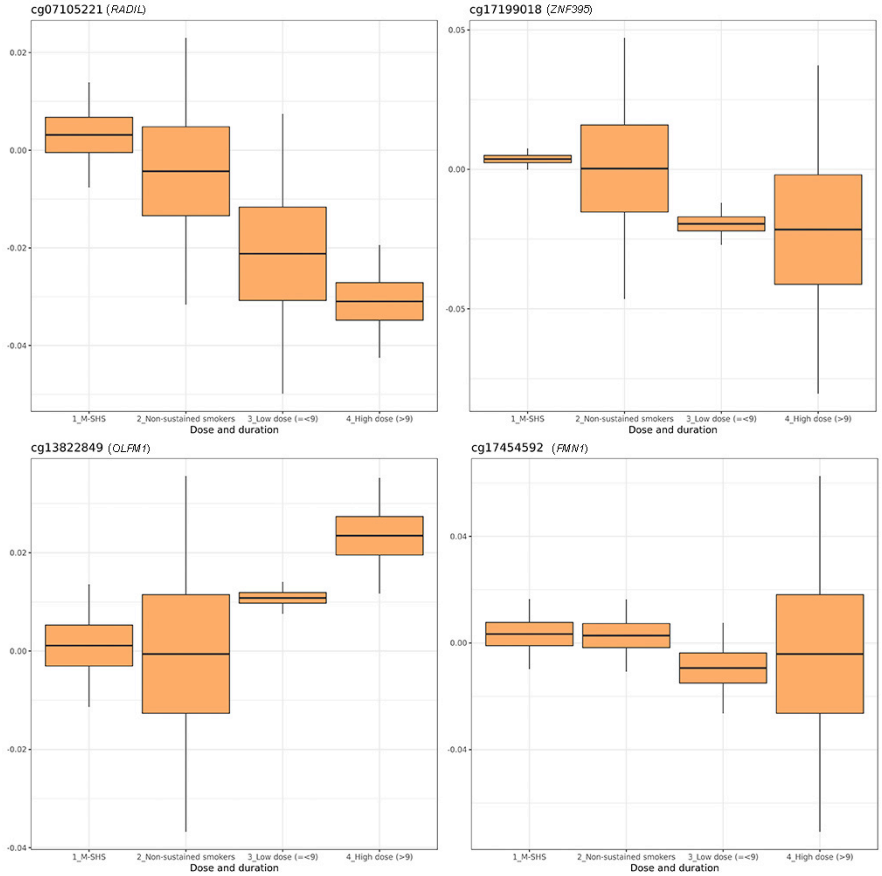

Figure S21. Boxplots showing the change of methylation in relation to control group (non-exposed mothers) in CpGs located near *RADIL, ZNF395, OLFM1* and *FMN1*. Sample size: 0_non-exposed (N=582); 1_M-SHS (N=351); 2_Non-sustained smokers (N=58); 3_Low dose (=<9 cig/d) (N=89); High dose (>9 cig/d) (N=15).

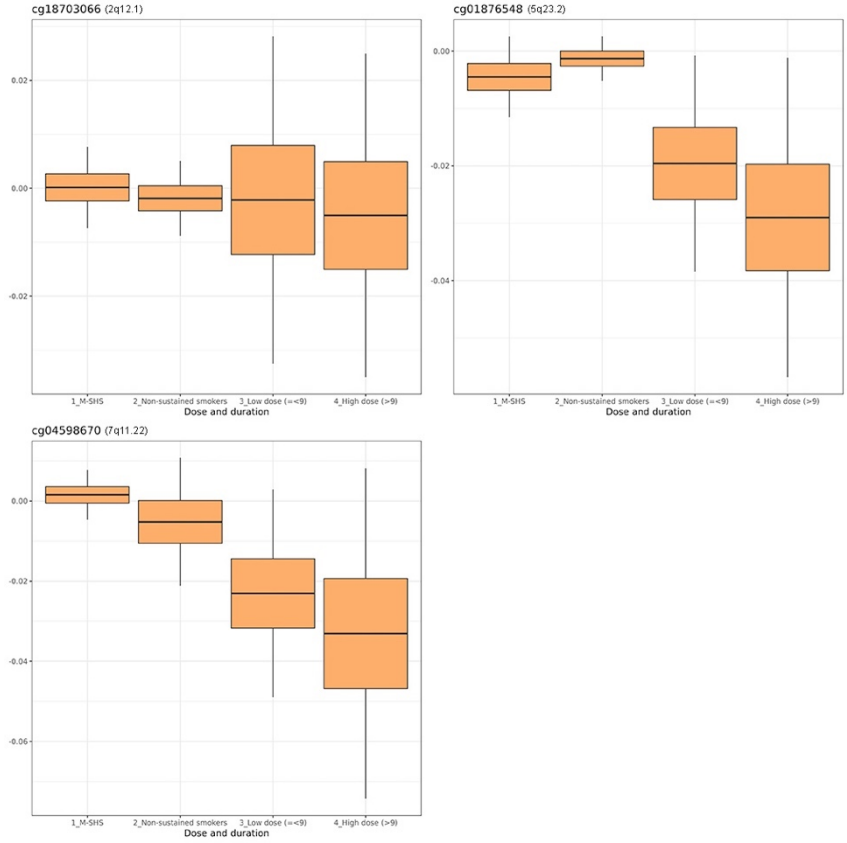

Figure S22. Boxplots showing the change of methylation in relation to control group (non-exposed mothers) in CpGs located at *2q12.1*, *5q23.2* and *7q11.22*. Sample size: 0_non-exposed (N=582); 1_M-SHS (N=351); 2_Non-sustained smokers (N=58); 3_Low dose (=<9 cig/d) (N=89); High dose (>9 cig/d) (N=15).

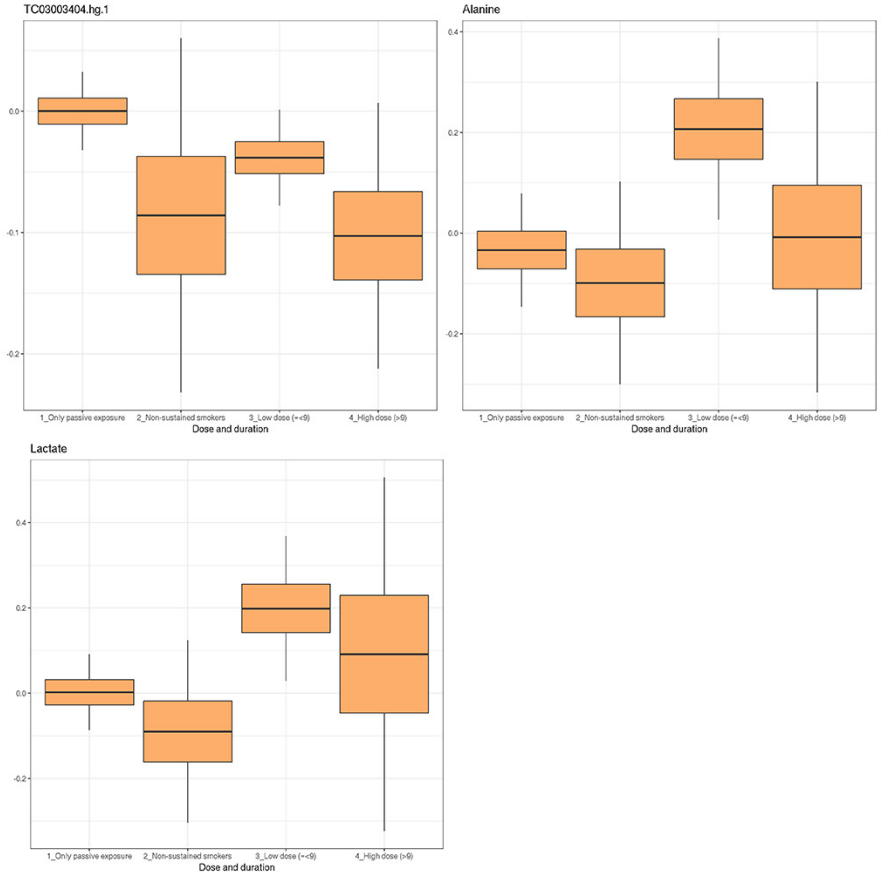

Figure S23. Boxplots showing the change of levels of a gene and two urinary metabolites in relation to control group (non-exposed mothers).

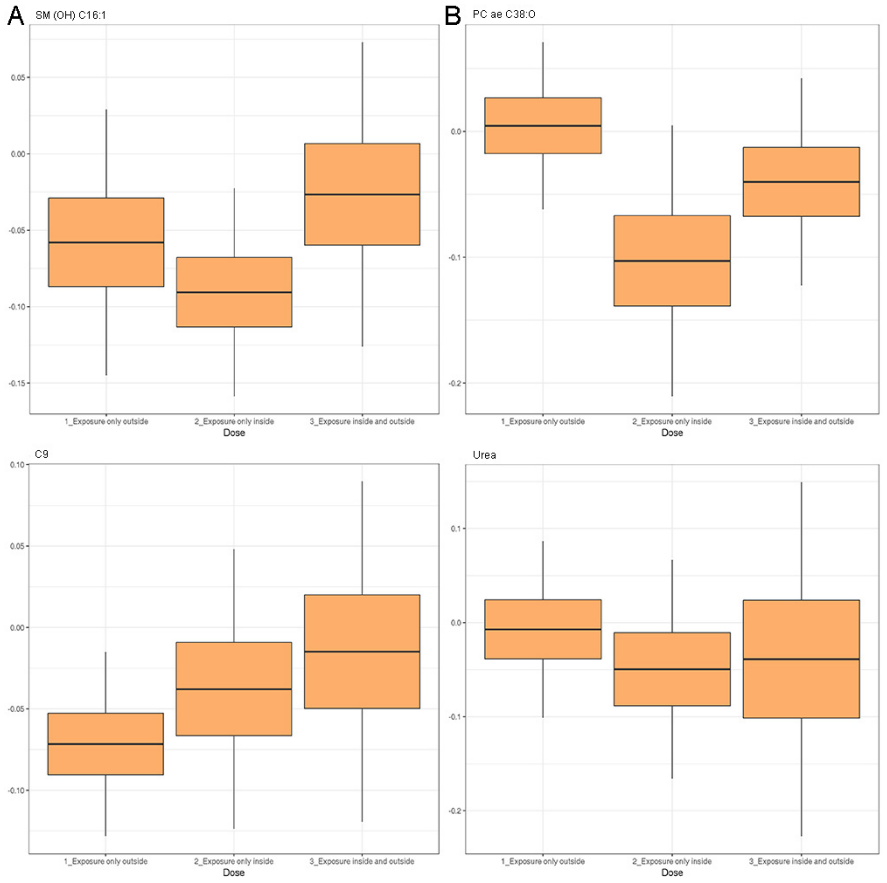

Figure S24 Boxplots showing the change of metabolites levels in whole population in relation to control group (non-exposed mothers). (A) Metabolites with lower levels in relation to G-SHS; (B) Metabolites with lower levels in children with detectable cotinine levels.

Table S1. Proteins targeted in each of the three Magnetic Human Luminex kits from Life Technologies.

Cytokines 30-plex		Apolipoprotein 5-plex	Adipokine 15-plex
FGF	EGF	ApoA1	IL-1Beta
IL1-Beta	IL-5	ApoB	IL-6
G-CSF	HGF	ApoE	IL-8
IL-10	VEGF	Adiponectin	IL-10
IL-13	IFN-Gamma	CRP	MCP-1
IL-6	IFN-Alfa		Leptin
IL-12	IL-1RA		SAA
RANTES	TNF-Alfa		HGF
Eotaxin	IL-2		Insulin
IL-17	IL-7		Lipocalin-2
MIP-Alfa	IP-10		TNF-Alfa
GM-CSF	IL-2R		BAFF
MIP-1Beta	MIG		Resistin
MCP-1	IL-4		C-Peptide
IL-15	IL-8		Serpine/PAI-1

Table S2. List of serum metabolites, class and the detection assay (Liquid Chromatography (LC) or Flow-Injection Analysis (FIA)).

Name	Class	Detection assay
Ala	Aminoacids	LC
Arg	Aminoacids	LC
Asn	Aminoacids	LC
Asp	Aminoacids	LC
Cit	Aminoacids	LC
Gln	Aminoacids	LC
Glu	Aminoacids	LC
Gly	Aminoacids	LC
His	Aminoacids	LC
Ile	Aminoacids	LC
Leu	Aminoacids	LC
Lys	Aminoacids	LC
Met	Aminoacids	LC
Orn	Aminoacids	LC
Phe	Aminoacids	LC
Pro	Aminoacids	LC

Ser	Aminoacids	LC
Thr	Aminoacids	LC
Trp	Aminoacids	LC
Tyr	Aminoacids	LC
Val	Aminoacids	LC
Ac-Orn	Biogenic amines	LC
ADMA	Biogenic amines	LC
Alpha-AAA	Biogenic amines	LC
Creatinine	Biogenic amines	LC
Histamine	Biogenic amines	LC
Kynunerine	Biogenic amines	LC
Met-SO	Biogenic amines	LC
Putrescine	Biogenic amines	LC
SDMA	Biogenic amines	LC
Serotonin	Biogenic amines	LC
Spermidine	Biogenic amines	LC
T4-OH-Pro	Biogenic amines	LC
Taurine	Biogenic amines	LC
Total DMA	Biogenic amines	LC
C0	Acylcarnitines	FIA
C10	Acylcarnitines	FIA
C10:1	Acylcarnitines	FIA
C10:2	Acylcarnitines	FIA
C12	Acylcarnitines	FIA
C12-DC	Acylcarnitines	FIA
C12:1	Acylcarnitines	FIA
C14	Acylcarnitines	FIA
C14:1	Acylcarnitines	FIA
C14:1-OH	Acylcarnitines	FIA
C14:2	Acylcarnitines	FIA
C14:2-OH	Acylcarnitines	FIA
C16	Acylcarnitines	FIA
C16-OH	Acylcarnitines	FIA
C16:1	Acylcarnitines	FIA
C16:1-OH	Acylcarnitines	FIA
C16:2	Acylcarnitines	FIA
C14:6-OH	Acylcarnitines	FIA
C18	Acylcarnitines	FIA
C18:1	Acylcarnitines	FIA

C18:1-OH	Acylcarnitines	FIA
C18:2	Acylcarnitines	FIA
C2	Acylcarnitines	FIA
C3	Acylcarnitines	FIA
C3-DC (C4-OH)	Acylcarnitines	FIA
C3-OH	Acylcarnitines	FIA
C3:1	Acylcarnitines	FIA
C4	Acylcarnitines	FIA
C4:1	Acylcarnitines	FIA
C6 (C4:1:DC)	Acylcarnitines	FIA
C5	Acylcarnitines	FIA
C5-M-DC	Acylcarnitines	FIA
C5-OH (C3-DC-M)	Acylcarnitines	FIA
C5:1	Acylcarnitines	FIA
C5:1-DC	Acylcarnitines	FIA
C5-DC (C6-OH)	Acylcarnitines	FIA
C6:1	Acylcarnitines	FIA
C7-DC	Acylcarnitines	FIA
C8	Acylcarnitines	FIA
C9	Acylcarnitines	FIA
lysoPC a C14:0	Glycerophospholipids	FIA
lysoPC a C16:0	Glycerophospholipids	FIA
lysoPC a C16:1	Glycerophospholipids	FIA
lysoPC a C17:0	Glycerophospholipids	FIA
lysoPC a C18:0	Glycerophospholipids	FIA
lysoPC a C18:1	Glycerophospholipids	FIA
lysoPC a C18:2	Glycerophospholipids	FIA
lysoPC a C20:3	Glycerophospholipids	FIA
lysoPC a C20:4	Glycerophospholipids	FIA
lysoPC a C24:0	Glycerophospholipids	FIA
lysoPC a C26:0	Glycerophospholipids	FIA
lysoPC a C28:0	Glycerophospholipids	FIA
lysoPC a C28:1	Glycerophospholipids	FIA
PC aa C24:0	Glycerophospholipids	FIA
PC aa C26:0	Glycerophospholipids	FIA
PC aa C28:1	Glycerophospholipids	FIA
PC aa C30:0	Glycerophospholipids	FIA
PC aa C30:2	Glycerophospholipids	FIA
PC aa C24:0	Glycerophospholipids	FIA

PC aa C26:0	Glycerophospholipids	FIA
PC aa C28:1	Glycerophospholipids	FIA
PC aa C30:0	Glycerophospholipids	FIA
PC aa C30:2	Glycerophospholipids	FIA
PC aa C32:0	Glycerophospholipids	FIA
PC aa C32:1	Glycerophospholipids	FIA
PC aa C32:2	Glycerophospholipids	FIA
PC aa C32:3	Glycerophospholipids	FIA
PC aa C34:1	Glycerophospholipids	FIA
PC aa C34:2	Glycerophospholipids	FIA
PC aa C34:3	Glycerophospholipids	FIA
PC aa C34:4	Glycerophospholipids	FIA
PC aa C36:0	Glycerophospholipids	FIA
PC aa C36:1	Glycerophospholipids	FIA
PC aa C36:2	Glycerophospholipids	FIA
PC aa C36:3	Glycerophospholipids	FIA
PC aa C36:4	Glycerophospholipids	FIA
PC aa C36:5	Glycerophospholipids	FIA
PC aa C36:6	Glycerophospholipids	FIA
PC aa C36:1	Glycerophospholipids	FIA
PC aa C38:0	Glycerophospholipids	FIA
PC aa C38:1	Glycerophospholipids	FIA
PC aa C38:3	Glycerophospholipids	FIA
PC aa C38:4	Glycerophospholipids	FIA
PC aa C38:5	Glycerophospholipids	FIA
PC aa C38:6	Glycerophospholipids	FIA
PC aa C40:1	Glycerophospholipids	FIA
PC aa C40:2	Glycerophospholipids	FIA
PC aa C40:3	Glycerophospholipids	FIA
PC aa C40:4	Glycerophospholipids	FIA
PC aa C40:5	Glycerophospholipids	FIA
PC aa C40:6	Glycerophospholipids	FIA
PC aa C42:0	Glycerophospholipids	FIA
PC aa C42:1	Glycerophospholipids	FIA
PC aa C42:2	Glycerophospholipids	FIA
PC aa C42:4	Glycerophospholipids	FIA
PC aa C42:5	Glycerophospholipids	FIA
PC aa C42:6	Glycerophospholipids	FIA
PC ae C30:0	Glycerophospholipids	FIA

PC ae C30:1	Glycerophospholipids	FIA
PC ae C30:2	Glycerophospholipids	FIA
PC ae C32:1	Glycerophospholipids	FIA
PC ae C32:2	Glycerophospholipids	FIA
PC ae C34:0	Glycerophospholipids	FIA
PC ae C34:1	Glycerophospholipids	FIA
PC ae C34:2	Glycerophospholipids	FIA
PC ae C34:3	Glycerophospholipids	FIA
PC ae C36:0	Glycerophospholipids	FIA
PC ae C36:1	Glycerophospholipids	FIA
PC ae C36:2	Glycerophospholipids	FIA
PC ae C36:3	Glycerophospholipids	FIA
PC ae C36:4	Glycerophospholipids	FIA
PC ae C36:5	Glycerophospholipids	FIA
PC ae C38:0	Glycerophospholipids	FIA
PC ae C38:1	Glycerophospholipids	FIA
PC ae C38:2	Glycerophospholipids	FIA
PC ae C38:3	Glycerophospholipids	FIA
PC ae C38:4	Glycerophospholipids	FIA
PC ae C38:5	Glycerophospholipids	FIA
PC ae C38:6	Glycerophospholipids	FIA
PC ae C40:1	Glycerophospholipids	FIA
PC ae C40:2	Glycerophospholipids	FIA
PC ae C40:3	Glycerophospholipids	FIA
PC ae C40:4	Glycerophospholipids	FIA
PC ae C40:5	Glycerophospholipids	FIA
PC ae C40:6	Glycerophospholipids	FIA
PC ae C42:1	Glycerophospholipids	FIA
PC ae C42:2	Glycerophospholipids	FIA
PC ae C42:3	Glycerophospholipids	FIA
PC ae C42:4	Glycerophospholipids	FIA
PC ae C42:5	Glycerophospholipids	FIA
PC ae C44:3	Glycerophospholipids	FIA
PC ae C44:4	Glycerophospholipids	FIA
PC ae C44:5	Glycerophospholipids	FIA
PC ae C44:6	Glycerophospholipids	FIA
SM (OH) C14:1	Sphingolipids	FIA
SM (OH) C16:1	Sphingolipids	FIA
SM (OH) C22:1	Sphingolipids	FIA

SM (OH) C22:2	Sphingolipids	FIA
SM (OH) C24:1	Sphingolipids	FIA
SM C16:0	Sphingolipids	FIA
SM C16:1	Sphingolipids	FIA
SM C18:0	Sphingolipids	FIA

a: acyl; aa: diacyl; ae: acyl-alkyl; lysoPC: lysophosphatidylcholine; PC: phosphatidylcholine; SM sphingomyelin; SM (OH): hydroxysphingomyelin.

Table S3. List of urine metabolites and their peak annotation and assignment.

Name	Annotation/Assignment
Leucine	STOCSY
Isoleucine	STOCSY
Valine	Chenomx, STOCSY
3-hydroxyisobutyrate	Chenomx
4-deoxyerythronic acid	Tentative
4-deoxythreonic acid	Tentative
Lactate	STOCSY
2-hydroxyisobutyrate	Chenomx
Alanine	Chenomx, STOCSY
Lysine	STOCSY
Acetate	2D
P-cresol sulfate	STOCSY
Succinate	Chenomx
Glutamine	STOCSY
Citrate	STOCSY
Dimethylamine	Chenomx, STOCSY
Trimethylamine	2D
Carnitine	STOCSY
Trimethylamine oxide	Chenomx, STOCSY
Scyllo-inositol	Tentative
Taurine	Chenomx, STOCSY
Glycine	Chenomx, STOCSY
Creatine	Chenomx, STOCSY
Creatinine	Chenomx, STOCSY
Hippurate	Chenomx, STOCSY
Formate	Chenomx, STOCSY
Pantothenic acid	Tentative
3-hydroxybutyrate/3-aminoisobytyrate	Tentative
3-hydroxyisovalerate	Chenomx, STOCSY

N-acetyl neuraminic acid	Tentative
Acetone	2D
5-oxoproline	2D
3-amonoisobutyrate	Tentative
Proline betaine	Tentative
Glucose	STOCSY
Sucrose	Chenomx, STOCSY
Urea	Chenomx
N-methyl-2-pyridone-5-carboxamide	STOCSY
p-hydroxyphenylacetate	Chenomx, STOCSY
Tyrosine	2D
3-Indoxylsulfate	Chenomx, STOCSY
N-methylpicolinic acid	Tentative
N-methylnicotinic acid	Chenomx, STOCSY
N1-methyl-nicotinamide	Chenomx, STOCSY

Chenomx: indicates good agreement with data obtained using alterative data processing workflow ("Chenomx"). Pearson correlation coefficient $R > 0.8$; STOCSY: indicates correlations observed with another resonance in the spectra that are either structurally od biochemically related; matching evidence on a representative INMA child sample.

CONTRIBUTIONS

This thesis has been done in collaboration with different research groups and laboratories which have contributed with data and analysis. Although these contributions have been mentioned along the text I would like to clarify the main contributions I have made.

1. Maternal tobacco smoking during pregnancy and placental DNA methylation, and its link with reproductive outcomes
    a. Statistical analysis in INMA cohort, both genome-wide differential DNA methylation analyses and association analyses between DNA methylation and reproductive outcomes.
    b. Quality control of results data from the rest of the cohorts
    c. Meta-analysis, both genome-wide differential DNA methylation analyses and association analyses between DNA methylation and reproductive outcomes.
    d. CpG site annotation
    e. Enrichment analyses
    f. Overlap of MSDP-sensitive CpG sites and birth weight loci
    g. Comparison of MSDP-sensitive CpGs in placenta with cord blood
    h. Generation of most of the plots

2. Exposure to prenatal and postnatal tobacco smoking and molecular signatures in children
    a. Harmonization of tobacco smoking variables
    b. Participation in protein selection
    c. RNA extraction
    d. Microarray processing of miRNAs
    e. Participation in quality control processes: RNA quality, miRNAs and gene expression
    f. All statistical analyses
    g. Generation of all plots

REFERENCES

1. Fda & CTP. Guidance for Industry Reporting Harmful and Potentially Harmful Constituents in Tobacco Products and Tobacco Smoke Under Section 904(a)(3) of the Federal Food, Drug, and Cosmetic Act. (2012).
2. Talhout, R. *et al.* Hazardous Compounds in Tobacco Smoke. *Int. J. Environ. Res. Public Heal. Int. J. Environ. Res. Public Heal. E-Mails Jan.van.Benthem@rivm.nl (J.v.B.); Piet.Wester@rivm.nl (P.W.); Antoon.Opperhuizen@rivm.nl* **83390,** 613–628 (1990).
3. Hatsukami, D. K., Stead, L. F. & Gupta, P. C. Tobacco Addiction: Diagnosis and Treatment. doi:10.1016/S0140-6736(08)60871-5
4. Noar, S. M. *et al.* Identifying principles for effective messages about chemicals in cigarette smoke. *Prev. Med. (Baltim).* **106,** 31–37 (2018).
5. (US), C. for D. C. and P., (US), N. C. for C. D. P. and H. P. & (US), O. on S. and H. Chemistry and Toxicology of Cigarette Smoke and Biomarkers of Exposure and Harm. (2010).
6. Mishra, A. *et al.* Harmful effects of nicotine. *Indian J. Med. Paediatr. Oncol.* **36,** 24–31 (2015).
7. Hecht, S. S. & Hoffmann, D. Tobacco-specific nitrosamines, an important group of carcinogens in tobacco and tobacco smoke. *Carcinogenesis* **9,** 875–84 (1988).
8. Benowitz, N. L. & Fraiman, J. B. Cardiovascular effects of electronic cigarettes. *Nat. Rev. Cardiol.* **14,** 447–456 (2017).

9. WHO | WHO report finds dramatic increase in life-saving tobacco control policies in last decade. *WHO* (2017).
10. WHO | World Health Organization. *WHO* (2018).
11. Yanbaeva, D. G., Dentener, M. A., Creutzberg, E. C., Wesseling, G. & Wouters, E. F. M. Systemic Effects of Smoking*. doi:10.1378/chest.06-2179
12. (US), O. of the S. G. & (US), O. on S. and H. *The Health Consequences of Smoking*. *The Health Consequences of Smoking: A Report of the Surgeon General* (Centers for Disease Control and Prevention (US), 2004).
13. Jaakkola, M. S. *et al.* Worldwide burden of disease from exposure to second-hand smoke: a retrospective analysis of data from 192 countries. *Lancet* **377,** 139–146 (2011).
14. Oberg, M., Jaakkola, M. S., Woodward, A., Peruga, A. & Prüss-Ustün, A. Worldwide burden of disease from exposure to second-hand smoke: a retrospective analysis of data from 192 countries. *Lancet (London, England)* **377,** 139–46 (2011).
15. Al-Sayed, E. M. & Ibrahim, K. S. Second-hand tobacco smoke and children. *Toxicol. Ind. Health* **30,** 635–644 (2014).
16. Ilicali, Ö. C., Keleş, N., Dèger, K., Sàgun, Ö. F. & Güldíken, Y. Evaluation of the Effect of Passive Smoking on Otitis Media in Children by an Objective Method: Urinary Cotinine Analysis. *Laryngoscope* **111,** 163–167 (2001).
17. Chang, J. S. *et al.* Parental Smoking and the Risk of Childhood Leukemia. *Am. J. Epidemiol.* **163,** 1091–1100 (2006).
18. Davis, C. L. *et al.* Passive Smoke Exposure and Its Effects on Cognition, Sleep, and Health Outcomes in Overweight and

Obese Children. *Child. Obes.* **12,** 119–125 (2016).

19. Johansson, A., Halling, A. & Hermansson, G. Indoor and outdoor smoking: impact on children's health. *Eur. J. Public Health* **13,** 61–6 (2003).
20. General, S. –, Atlanta, G.] & Dept, U. S. The Health Consequences of Involuntary Exposure to Tobacco Smoke National Library of Medicine Cataloging in Publication. *Heal. Hum. Serv. Centers Dis. Control Prev. Centers Dis. Control Prev.* (2006).
21. Scientific Committee on Tobacco and Health (SCOTH) Secondhand Smoke: Review of evidence since 1998 Update of evidence on health effects of secondhand smoke.
22. Andres, R. L. & Day, M.-C. Perinatal complications associated with maternal tobacco use. *Semin. Neonatol.* **5,** 231–241 (2000).
23. Banderali, G. *et al.* Short and long term health effects of parental tobacco smoking during pregnancy and lactation: a descriptive review. *J. Transl. Med.* **13,** 327 (2015).
24. Yolton, K., Dietrich, K., Auinger, P., Lanphear, B. P. & Hornung, R. Exposure to environmental tobacco smoke and cognitive abilities among U.S. children and adolescents. *Environ. Health Perspect.* **113,** 98–103 (2005).
25. Mund, M., Louwen, F., Klingelhoefer, D. & Gerber, A. Smoking and pregnancy--a review on the first major environmental risk factor of the unborn. *Int. J. Environ. Res. Public Health* **10,** 6485–99 (2013).
26. Reese, S. E. *et al.* DNA Methylation Score as a Biomarker in

Newborns for Sustained Maternal Smoking during Pregnancy. *Environ. Health Perspect.* **125,** (2016).

27. Huang, S.-H. *et al.* The effects of maternal smoking exposure during pregnancy on postnatal outcomes: A cross sectional study. *J. Chinese Med. Assoc.* **80,** 796–802 (2017).
28. Smoking prevalence and cessation before and during pregnancy: data from the birth certificate - Cerca amb Google. Available at: https://www.google.es/search?q=Smoking+prevalence+and+cessation+before+and+during+pregnancy%3A+data+from+the+birth+certificate&oq=Smoking+prevalence+and+cessation+before+and+during+pregnancy%3A+data+from+the+birth+certificate&aqs=chrome..69i57j69i61.463j0. (Accessed: 16th January 2018)
29. Smedberg, J., Lupattelli, A., Mårdby, A.-C. & Nordeng, H. Characteristics of women who continue smoking during pregnancy: a cross-sectional study of pregnant women and new mothers in 15 European countries. *BMC Pregnancy Childbirth* **14,** 213 (2014).
30. Lange, S., Probst, C., Rehm, J. & Popova, S. National, regional, and global prevalence of smoking during pregnancy in the general population: a systematic review and meta-analysis. *Lancet. Glob. Heal.* **6,** e769–e776 (2018).
31. Soma-Pillay, P. *et al.* Physiological changes in pregnancy. *Cardiovasc. J. Afr.* **27,** 89–94 (2016).
32. Mian, M. F., Lauzon, N. M., Stämpfli, M. R., Mossman, K. L. & Ashkar, A. A. Impairment of human NK cell cytotoxic

activity and cytokine release by cigarette smoke. *J. Leukoc. Biol.* **83,** 774–784 (2008).

33. Sabra, S., Gratacós, E. & Gómez Roig, M. D. Smoking-Induced Changes in the Maternal Immune, Endocrine, and Metabolic Pathways and Their Impact on Fetal Growth: A Topical Review. *Fetal Diagn. Ther.* **41,** 241–250 (2017).
34. Fang, F., Luo, Z.-C., Dejemli, A., Delvin, E. & Zhang, J. Maternal Smoking and Metabolic Health Biomarkers in Newborns. *PLoS One* **10,** e0143660 (2015).
35. Jung, E. J. *et al.* Placental pathologic changes and perinatal outcomes in placenta previa. *Placenta* **63,** 15–20 (2018).
36. Ashfaq, M., Janjua, M. Z. & Nawaz, M. Effects of maternal smoking on placental morphology. *J. Ayub Med. Coll. Abbottabad* **15,** 12–5
37. Wei, J., Liu, C.-X., Gong, T.-T., Wu, Q.-J. & Wu, L. Cigarette smoking during pregnancy and preeclampsia risk: a systematic review and meta-analysis of prospective studies. *Oncotarget* **6,** 43667–78 (2015).
38. Karumanchi, S. A. & Levine, R. J. How does smoking reduce the risk of preeclampsia? *Hypertens. (Dallas, Tex. 1979)* **55,** 1100–1 (2010).
39. Perni, U. C., Wikstrom, A.-K., Cnattingius, S. & Villamor, E. Interpregnancy Change in Smoking Habits and Risk of Preeclampsia: A Population-Based Study. *Am. J. Hypertens.* **25,** 372–378 (2012).
40. Hammoud, A. O. *et al.* Smoking in pregnancy revisited: Findings from a large population-based study. *Am. J. Obstet.*

Gynecol. **192,** 1856–1862 (2005).

41. Stone, C. D., Diallo, O., Shyken, J. & Leet, T. The combined effect of maternal smoking and obesity on the risk of preeclampsia. *J. Perinat. Med.* **35,** 28–31 (2007).
42. Newman, M. G., Lindsay, M. K. & Graves, W. Cigarette smoking and pre-eclampsia: their association and effects on clinical outcomes. *J. Matern. Fetal. Med.* **10,** 166–70 (2001).
43. Zhang, J., Klebanoff, M. A., Levine, R. J., Puri, M. & Moyer, P. The puzzling association between smoking and hypertension during pregnancy. *Am. J. Obstet. Gynecol.* **181,** 1407–13 (1999).
44. Coonrod, D. V, Hickok, D. E., Zhu, K., Easterling, T. R. & Daling, J. R. Risk factors for preeclampsia in twin pregnancies: a population-based cohort study. *Obstet. Gynecol.* **85,** 645–50 (1995).
45. England, L. J. *et al.* Smoking before pregnancy and risk of gestational hypertension and preeclampsia. *Am. J. Obstet. Gynecol.* **186,** 1035–40 (2002).
46. Xiong, X. *et al.* Maternal smoking and preeclampsia. *J. Reprod. Med.* **45,** 727–32 (2000).
47. Ioka, A., Tsukuma, H. & Nakamuro, K. Lifestyles and pre-eclampsia with special attention to cigarette smoking. *J. Epidemiol.* **13,** 90–5 (2003).
48. Basso, O. *et al.* Subfecundity as a correlate of preeclampsia: a study within the Danish National Birth Cohort. *Am. J. Epidemiol.* **157,** 195–202 (2003).
49. Mortensen, J. T., Thulstrup, A. M., Larsen, H., Møller, M. &

Sørensen, H. T. Smoking, sex of the offspring, and risk of placental abruption, placenta previa, and preeclampsia: a population-based cohort study. *Acta Obstet. Gynecol. Scand.* **80,** 894–8 (2001).

50. Martin, C. L., Hall, M. H. & Campbell, D. M. The effect of smoking on pre-eclampsia in twin pregnancy. *BJOG* **107,** 745–9 (2000).
51. Odegård, R. A., Vatten, L. J., Nilsen, S. T., Salvesen, K. A. & Austgulen, R. Risk factors and clinical manifestations of pre-eclampsia. *BJOG* **107,** 1410–6 (2000).
52. Lindqvist, P. G. & Marsál, K. Moderate smoking during pregnancy is associated with a reduced risk of preeclampsia. *Acta Obstet. Gynecol. Scand.* **78,** 693–7 (1999).
53. Cnattingius, S., Mills, J. L., Yuen, J., Eriksson, O. & Salonen, H. The paradoxical effect of smoking in preeclamptic pregnancies: smoking reduces the incidence but increases the rates of perinatal mortality, abruptio placentae, and intrauterine growth restriction. *Am. J. Obstet. Gynecol.* **177,** 156–61 (1997).
54. Knuist, M., Bonsel, G. J., Zondervan, H. A. & Treffers, P. E. Risk factors for preeclampsia in nulliparous women in distinct ethnic groups: a prospective cohort study. *Obstet. Gynecol.* **92,** 174–8 (1998).
55. Lisonkova, S. & Joseph, K. S. Incidence of preeclampsia: risk factors and outcomes associated with early- versus late-onset disease. *Am. J. Obstet. Gynecol.* **209,** 544.e1-544.e12 (2013).
56. Wang, A., Rana, S. & Karumanchi, S. A. Preeclampsia: The

Role of Angiogenic Factors in Its Pathogenesis. *Physiology* **24,** 147–158 (2009).

57. Konukoglu, D. & Uzun, H. Endothelial Dysfunction and Hypertension. in *Advances in experimental medicine and biology* **956,** 511–540 (2016).
58. Dharmashankar, K. & Widlansky, M. E. Vascular endothelial function and hypertension: insights and directions. *Curr. Hypertens. Rep.* **12,** 448–55 (2010).
59. Ataga, K. I. *et al.* Albuminuria Is Associated with Endothelial Dysfunction and Elevated Plasma Endothelin-1 in Sickle Cell Anemia. *PLoS One* **11,** e0162652 (2016).
60. Paisley, K. E. *et al.* Endothelial dysfunction and inflammation in asymptomatic proteinuria. *Kidney Int.* **63,** 624–633 (2003).
61. Shah, N. R. & Bracken, M. B. A systematic review and meta-analysis of prospective studies on the association between maternal cigarette smoking and preterm delivery. *Am. J. Obstet. Gynecol.* **182,** 465–72 (2000).
62. Rogers, J. M. Tobacco and pregnancy. *Reprod. Toxicol.* **28,** 152–160 (2009).
63. Ion, R. & Bernal, A. L. Smoking and Preterm Birth. *Reprod. Sci.* **22,** 918–926 (2015).
64. Blencowe, H. *et al.* Born too soon: the global epidemiology of 15 million preterm births. *Reprod. Health* **10 Suppl 1,** S2 (2013).
65. Pineles, B. L., Hsu, S., Park, E. & Samet, J. M. Systematic Review and Meta-Analyses of Perinatal Death and Maternal Exposure to Tobacco Smoke During Pregnancy. *Am. J.*

Epidemiol. **184,** 87–97 (2016).

66. Mohammadi, S. *et al.* Toxic compounds from tobacco in placenta samples analyzed by UPLC-QTOF-MS. *J. Pharm. Biomed. Anal.* **145,** 331–338 (2017).
67. Gibbs, K., Collaco, J. M. & McGrath-Morrow, S. A. Impact of Tobacco Smoke and Nicotine Exposure on Lung Development. *Chest* **149,** 552–561 (2016).
68. Oken, E., Levitan, E. B. & Gillman, M. W. Maternal smoking during pregnancy and child overweight: systematic review and meta-analysis. *Int. J. Obes.* **32,** 201–210 (2008).
69. Burke, H. *et al.* Prenatal and Passive Smoke Exposure and Incidence of Asthma and Wheeze: Systematic Review and Meta-analysis. *Pediatrics* **129,** 735–744 (2012).
70. Hollams, E. M., de Klerk, N. H., Holt, P. G. & Sly, P. D. Persistent Effects of Maternal Smoking during Pregnancy on Lung Function and Asthma in Adolescents. *Am. J. Respir. Crit. Care Med.* **189,** 401–407 (2014).
71. Shea, A. & Steiner, M. Cigarette smoking during pregnancy. *Nicotine Tob. Res.* **10,** 267–278 (2008).
72. Wild, C. P. Complementing the genome with an ‘exposome’: The outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiol. Biomarkers Prev.* **14,** 1847–1850 (2005).
73. Kim, K.-N. & Hong, Y.-C. The exposome and the future of epidemiology: a vision and prospect. *Environ. Health Toxicol.* **32,** e2017009 (2017).
74. Peters, A., Hoek, G. & Katsouyanni, K. Understanding the link

between environmental exposures and health: does the exposome promise too much?: Figure 1. *J. Epidemiol. Community Health* **66,** 103–105 (2012).

75. Vrijheid, M. The exposome: a new paradigm to study the impact of environment on health. *Thorax* **69,** 876–878 (2014).
76. Bauer, T. *et al.* Environment-induced epigenetic reprogramming in genomic regulatory elements in smoking mothers and their children. *Mol. Syst. Biol.* **12,** 861 (2016).
77. Weinhold, B. Epigenetics: the science of change. *Environ. Health Perspect.* **114,** A160-7 (2006).
78. Bianco-Miotto, T., Craig, J. M., Gasser, Y. P., van Dijk, S. J. & Ozanne, S. E. Epigenetics and DOHaD: from basics to birth and beyond. *J. Dev. Orig. Health Dis.* **8,** 513–519 (2017).
79. Kim, G. H., Ryan, J. J., Marsboom, G. & Archer, S. L. Epigenetic mechanisms of pulmonary hypertension. *Pulm. Circ.* **1,** 347–56 (2011).
80. CpG island - an overview | ScienceDirect Topics. Available at: https://www.sciencedirect.com/topics/neuroscience/cpg-island. (Accessed: 12th July 2018)
81. Maccani, M. A. & Marsit, C. J. Epigenetics in the placenta. *Am. J. Reprod. Immunol.* **62,** 78–89 (2009).
82. Alsharafi, W. A., Xiao, B., Abuhamed, M. M. & Luo, Z. miRNAs: biological and clinical determinants in epilepsy. *Front. Mol. Neurosci.* **8,** 59 (2015).
83. Ernst, J. & Kellis, M. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat. Biotechnol.* **28,** 817–25 (2010).

84. Roadmap Epigenomics Project - Data. Available at: http://www.roadmapepigenomics.org/data/. (Accessed: 14th February 2018)
85. Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473,** 43–49 (2011).
86. Consortium, T. E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).
87. Frith, M. C., Pheasant, M. & Mattick, J. S. Genomics: The amazing complexity of the human transcriptome. *Eur. J. Hum. Genet.* **13,** 894–897 (2005).
88. Casamassimi, A., Federico, A., Rienzo, M., Esposito, S. & Ciccodicola, A. Transcriptome Profiling in Human Diseases: New Advances and Perspectives. *Int. J. Mol. Sci.* **18,** (2017).
89. The metabolome and metabolic reactions | EMBL-EBI Train online. Available at: https://www.ebi.ac.uk/training/online/course/introduction-metabolomics/what-metabolomics/metabolome-and-metabolic-reactions. (Accessed: 22nd May 2018)
90. Barrett, J. R. Programming the Future: Epigenetics in the Context of DOHaD. *Environ. Health Perspect.* **125,** A72 (2017).
91. Harding, R. & Bocking, A. D. *Fetal growth and development*. (Cambridge University Press, 2001).
92. Prenatal Development and The 3 Stages | Biology Dictionary. Available at: https://biologydictionary.net/prenatal-development/#prettyPhoto. (Accessed: 8th July 2018)
93. Fetal Development, Human - Biology Encyclopedia - cells,

body, function, process, system, chromosomes, blood, cycle. Available at: http://www.biologyreference.com/Ep-Fl/Fetal-Development-Human.html. (Accessed: 10th July 2018)

94. Slieker, R. C. *et al.* DNA Methylation Landscapes of Human Fetal Development. *PLoS Genet.* **11,** e1005583 (2015).

95. Ishida, M. & Moore, G. E. The role of imprinted genes in humans. *Mol. Aspects Med.* **34,** 826–840 (2013).

96. Louis, G. M. B., Smarr, M. M. & Patel, C. J. The Exposome Research Paradigm: An Opportunity to Understand the Environmental Basis for Human Health and Disease HHS Public Access. *Curr Env. Heal. Rep* **4,** 89–98 (2017).

97. Lokk, K. *et al.* DNA methylome profiling of human tissues identifies global and tissue-specific methylation patterns. *Genome Biol.* **15,** r54 (2014).

98. Mele, M. *et al.* The human transcriptome across tissues and individuals. *Science (80-.).* **348,** 660–665 (2015).

99. Components of blood (article) | Khan Academy. Available at: https://www.khanacademy.org/science/biology/human-biology/circulatory-pulmonary/a/components-of-the-blood. (Accessed: 12th July 2018)

100. Bauer, M. *et al.* A varying T cell subtype explains apparent tobacco smoking induced single CpG hypomethylation in whole blood. *Clin. Epigenetics* **7,** 81 (2015).

101. Bauer, M. Cell-type-specific disturbance of DNA methylation pattern: a chance to get more benefit from and to minimize cohorts for epigenome-wide association studies. *Int. J. Epidemiol.* **47,** 917–927 (2018).

102. Schuyler, R. P. *et al.* Distinct Trends of DNA Methylation Patterning in the Innate and Adaptive Immune Systems. *Cell Rep.* **17,** 2101–2111 (2016).
103. Teschendorff, A. E. & Relton, C. L. Statistical and integrative system-level analysis of DNA methylation data. *Nat. Rev. Genet.* (2017). doi:10.1038/nrg.2017.86
104. Gude, N. M., Roberts, C. T., Kalionis, B. & King, R. G. Growth and function of the normal human placenta. *Thromb. Res.* **114,** 397–407 (2004).
105. Huppertz, B. The anatomy of the normal placenta. *J. Clin. Pathol.* **61,** 1296–1302 (2008).
106. Burton, G. J. & Fowden, A. L. The placenta: a multifaceted, transient organ. doi:10.1098/rstb.2014.0066
107. Heger, A., Sator, M., Walch, K. & Pietrowski, D. Smoking Decreases Endometrial Thickness in IVF/ICSI Patients. *Geburtshilfe Frauenheilkd.* **78,** 78–82 (2018).
108. Anatomy Of Placenta Family GP – Lifeinharmony. Available at: http://www.lifeinharmony.me/anatomy-of-placenta/anatomy-of-placenta-family-gp. (Accessed: 26th May 2018)
109. Placenta - What Is It and How It Works | Biology Dictionary. Available at: https://biologydictionary.net/placenta/#prettyPhoto. (Accessed: 26th May 2018)
110. Grigoriu, A. *et al.* Cell specific patterns of methylation in the human placenta. *Epigenetics* **6,** 368–79 (2011).
111. Salhab, A. *et al.* Partially methylated domains are hallmarks

of a cell specific epigenome topology. (2018). doi:10.1101/249334

112. Schroeder, D. I. *et al.* The human placenta methylome. doi:10.1073/pnas.1215145110

113. Cassidy, F. C. & Charalambous, M. Genomic imprinting, growth and maternal–fetal interactions. doi:10.1242/jeb.164517

114. Breton, C. V. *et al.* Small-Magnitude Effect Sizes in Epigenetic End Points are Important in Children's Environmental Health Studies: The Children's Environmental Health and Disease Prevention Research Center's Epigenetics Working Group. *Environ. Health Perspect.* **125,** 511–526 (2017).

115. Breitling, L. P., Yang, R., Korn, B., Burwinkel, B. & Brenner, H. Tobacco-Smoking-Related Differential DNA Methylation: 27K Discovery and Replication. *Am. J. Hum. Genet.* **88,** 450–457 (2011).

116. Zeilinger, S. *et al.* Tobacco Smoking Leads to Extensive Genome-Wide Changes in DNA Methylation. *PLoS One* **8,** e63812 (2013).

117. Harlid, S., Xu, Z., Panduri, V., Sandler, D. P. & Taylor, J. A. CpG Sites Associated with Cigarette Smoking: Analysis of Epigenome-Wide Data from the Sister Study. *Environ. Health Perspect.* **122,** 673–8 (2014).

118. Tsaprouni, L. G. *et al.* Cigarette smoking reduces DNA methylation levels at multiple genomic loci but the effect is partially reversible upon cessation. *Epigenetics* **9,** 1382–1396

(2014).

119. Guida, F. *et al.* Dynamics of smoking-induced genome-wide methylation changes with time since smoking cessation. *Hum. Mol. Genet.* **24,** 2349–2359 (2015).

120. Zaghlool, S. B. *et al.* Association of DNA methylation with age, gender, and smoking in an Arab population. *Clin. Epigenetics* **7,** 6 (2015).

121. Philibert, R. A., Beach, S. R. H. & Brody, G. H. Demethylation of the aryl hydrocarbon receptor repressor as a biomarker for nascent smokers. *Epigenetics* **7,** 1331–1338 (2012).

122. Wan, E. S. *et al.* Cigarette smoking behaviors and time since quitting are associated with differential DNA methylation across the human genome. *Hum. Mol. Genet.* **21,** 3073–3082 (2012).

123. Philibert, R. A., Beach, S. R. H., Lei, M.-K. & Brody, G. H. Changes in DNA methylation at the aryl hydrocarbon receptor repressor may be a new biomarker for smoking. *Clin. Epigenetics* **5,** 19 (2013).

124. Shenker, N. S. *et al.* Epigenome-wide association study in the European Prospective Investigation into Cancer and Nutrition (EPIC-Turin) identifies novel genetic loci associated with smoking. *Hum. Mol. Genet.* **22,** 843–851 (2013).

125. Sun, Y. V. *et al.* Epigenomic association analysis identifies smoking-related DNA methylation sites in African Americans. *Hum. Genet.* **132,** 1027–1037 (2013).

126. Besingi, W. & Johansson, Å. Smoke-related DNA methylation

changes in the etiology of human disease. *Hum. Mol. Genet.* **23,** 2290–2297 (2014).

127. Dogan, M. V *et al.* The effect of smoking on DNA methylation of peripheral blood mononuclear cells from African American women. *BMC Genomics* **15,** 151 (2014).
128. Elliott, H. R. *et al.* Differences in smoking associated DNA methylation patterns in South Asians and Europeans. *Clin. Epigenetics* **6,** 4 (2014).
129. Shenker, N. S. *et al.* DNA Methylation as a Long-term Biomarker of Exposure to Tobacco Smoke. *Epidemiology* **24,** 712–716 (2013).
130. Zhang, Y., Yang, R., Burwinkel, B., Breitling, L. P. & Brenner, H. F2RL3 Methylation as a Biomarker of Current and Lifetime Smoking Exposures. *Environ. Health Perspect.* **122,** 131–7 (2013).
131. Breitling, L. P., Salzmann, K., Rothenbacher, D., Burwinkel, B. & Brenner, H. Smoking, F2RL3 methylation, and prognosis in stable coronary heart disease. *Eur. Heart J.* **33,** 2841–2848 (2012).
132. Gao, X., Jia, M., Zhang, Y., Breitling, L. P. & Brenner, H. DNA methylation changes of whole blood cells in response to active smoking exposure in adults: a systematic review of DNA methylation studies. *Clin. Epigenetics* **7,** 113 (2015).
133. Joehanes, R. *et al.* Epigenetic Signatures of Cigarette Smoking. *Circ. Cardiovasc. Genet.* **9,** 436–447 (2016).
134. Zhu, X. *et al.* Genome-Wide Analysis of DNA Methylation and Cigarette Smoking in a Chinese Population. *Environ.*

Health Perspect. **124,** 966–73 (2016).

135. Lee, M. K., Hong, Y., Kim, S.-Y., London, S. J. & Kim, W. J. DNA methylation and smoking in Korean adults: epigenome-wide association study. *Clin. Epigenetics* **8,** 103 (2016).
136. Gonseth, S. *et al.* Genetic contribution to variation in DNA methylation at maternal smoking-sensitive loci in exposed neonates. *Epigenetics* **11,** 664–673 (2016).
137. Reynolds, L. M. *et al.* Secondhand Tobacco Smoke Exposure Associations with DNA Methylation of the Aryl Hydrocarbon Receptor Repressor. *Nicotine Tob. Res.* **19,** ntw219 (2016).
138. Joubert, B. R. *et al.* 450K Epigenome-Wide Scan Identifies Differential DNA Methylation in Newborns Related to Maternal Smoking during Pregnancy. *Environ. Health Perspect.* **120,** 1425–1431 (2012).
139. Küpers, L. K. *et al.* DNA methylation mediates the effect of maternal smoking during pregnancy on birthweight of the offspring. *Int. J. Epidemiol.* **44,** 1224–1237 (2015).
140. Richmond, R. C. *et al.* Prenatal exposure to maternal smoking and offspring DNA methylation across the lifecourse: findings from the Avon Longitudinal Study of Parents and Children (ALSPAC). *Hum. Mol. Genet.* **24,** 2201–2217 (2015).
141. Joubert, B. R. *et al.* DNA Methylation in Newborns and Maternal Smoking in Pregnancy: Genome-wide Consortium Meta-analysis. *Am. J. Hum. Genet.* **98,** 680–696 (2016).
142. Markunas, C. A. *et al.* Identification of DNA Methylation Changes in Newborns Related to Maternal Smoking during Pregnancy. *Environ. Health Perspect.* **122,** 1147–53 (2014).

143. Breton, C. V. *et al.* Prenatal Tobacco Smoke Exposure Is Associated with Childhood DNA CpG Methylation. *PLoS One* **9,** e99716 (2014).
144. Flom, J. D. *et al.* Prenatal Smoke Exposure and Genomic DNA Methylation in a Multiethnic Birth Cohort. *Cancer Epidemiol. Biomarkers Prev.* **20,** 2518–2523 (2011).
145. Miyake, K. *et al.* Association between DNA methylation in cord blood and maternal smoking: The Hokkaido Study on Environment and Children's Health. *Sci. Rep.* **8,** 5654 (2018).
146. Witt, S. H. *et al.* Impact on birth weight of maternal smoking throughout pregnancy mediated by DNA methylation. *BMC Genomics* **19,** 290 (2018).
147. Maccani, J. Z. J. & Maccani, M. Altered placental DNA methylation patterns associated with maternal smoking: current perspectives. *Adv. Genomics Genet.* **2015,** 205 (2015).
148. Tehranifar, P. *et al.* Maternal cigarette smoking during pregnancy and offspring DNA methylation in midlife. *Epigenetics* **13,** 129–134 (2018).
149. Health, N. C. for C. D. P. and H. P. (US) O. on S. and. *The Health Consequences of Smoking—50 Years of Progress*. *The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General* (Centers for Disease Control and Prevention (US), 2014).
150. Novakovic, B. *et al.* Postnatal stability, tissue, and time specific effects of *AHRR* methylation change in response to maternal smoking in pregnancy. *Epigenetics* **9,** 377–386 (2014).

151. Morales, E. *et al.* Genome-wide DNA methylation study in human placenta identifies novel loci associated with maternal smoking during pregnancy. *Int. J. Epidemiol.* 1644–1655 (2016). doi:10.1093/ije/dyw196

152. Suter, M. *et al.* In utero tobacco exposure epigenetically modifies placental CYP1A1 expression. *Metabolism* **59,** 1481–1490 (2010).

153. Suter, M. *et al.* Maternal tobacco use modestly alters correlated epigenome-wide placental DNA methylation and gene expression. *Epigenetics* **6,** 1284–1294 (2011).

154. Wilhelm-Benartzi, C. S. *et al.* In utero exposures, infant growth, and DNA methylation of repetitive elements and developmentally related genes in human placenta. *Environ. Health Perspect.* **120,** 296–302 (2012).

155. Appleton, A. A. *et al.* Patterning in Placental 11-B Hydroxysteroid Dehydrogenase Methylation According to Prenatal Socioeconomic Adversity. *PLoS One* **8,** e74691 (2013).

156. Maccani, J. Z., Koestler, D. C., Houseman, E. A., Marsit, C. J. & Kelsey, K. T. Placental DNA methylation alterations associated with maternal tobacco smoking at the *RUNX3* gene are also associated with gestational age. *Epigenomics* **5,** 619–630 (2013).

157. Paquette, A. G. *et al.* Placental HTR2A methylation is associated with infant neurobehavioral outcomes. *Epigenetics* **8,** 796–801 (2013).

158. Chhabra, D. *et al.* Fetal lung and placental methylation is

associated with in utero nicotine exposure. *Epigenetics* **9,** 1473–1484 (2014).

159. Stroud, L. R. *et al.* Epigenetic Regulation of Placental *NR3C1* : Mechanism Underlying Prenatal Programming of Infant Neurobehavior by Maternal Smoking? *Child Dev.* **87,** 49–60 (2016).
160. Marczylo, E. L., Amoako, A. A., Konje, J. C., Gant, T. W. & Marczylo, T. H. Smoking induces differential miRNA expression in human spermatozoa: A potential transgenerational epigenetic concern? *Epigenetics* **7,** 432–439 (2012).
161. Wang, G. *et al.* Persistence of Smoking-Induced Dysregulation of MiRNA Expression in the Small Airway Epithelium Despite Smoking Cessation. *PLoS One* **10,** e0120824 (2015).
162. Opstad, T. B., Brusletto, B. S., Arnesen, H., Pettersen, A. Å. & Seljeflot, I. Cigarette smoking represses expression of cytokine IL-12 and its regulator miR-21—An observational study in patients with coronary artery disease. *Immunobiology* **222,** 169–175 (2017).
163. Willinger, C. M. *et al.* MicroRNA Signature of Cigarette Smoking and Evidence for a Putative Causal Role of MicroRNAs in Smoking-Related Inflammation and Target Organ DamageCLINICAL PERSPECTIVE. *Circ. Cardiovasc. Genet.* **10,** e001678 (2017).
164. Civelek, M. A Smoking-Associated miRNA-mRNA Coexpression Network. *Circ. Cardiovasc. Genet.* **10,** e001914

(2017).

165. Maccani, M. A. *et al.* Maternal cigarette smoking during pregnancy is associated with downregulation of miR-16, miR-21, and miR-146a in the placenta. *Epigenetics* **5,** 583–9 (2010).

166. Herberth, G. *et al.* Maternal and cord blood miR-223 expression associates with prenatal tobacco smoke exposure and low regulatory T-cell numbers. *J. Allergy Clin. Immunol.* **133,** 543–550.e4 (2014).

167. Zhang, L. *et al.* Impact of Smoking Cessation on Global Gene Expression in the Bronchial Epithelium of Chronic Smokers. *Cancer Prev. Res.* **1,** 112–118 (2008).

168. Harvey, B.-G. *et al.* Modification of gene expression of the small airway epithelium in response to cigarette smoking. *J. Mol. Med.* **85,** 39–53 (2006).

169. Chari, R. *et al.* Effect of active smoking on the human bronchial epithelium transcriptome. *BMC Genomics* **8,** 297 (2007).

170. Sridhar, S. *et al.* Smoking-induced gene expression changes in the bronchial airway are reflected in nasal and buccal epithelium. *BMC Genomics* **9,** 259 (2008).

171. Zhang, X. *et al.* Similarities and differences between smoking-related gene expression in nasal and bronchial epithelium. *Physiol. Genomics* **41,** 1–8 (2010).

172. Imkamp, K. *et al.* Nasal epithelium as a proxy for bronchial epithelium for smoking-induced gene expression and expression Quantitative Trait Loci. *J. Allergy Clin. Immunol.*

(2018). doi:10.1016/j.jaci.2018.01.047

173. Beineke, P. *et al.* A whole blood gene expression-based signature for smoking status. *BMC Med. Genomics* **5,** 58 (2012).

174. Huan, T. *et al.* A Whole-Blood Transcriptome Meta-Analysis Identifies Gene Expression Signatures of Cigarette Smoking. *Hum. Mol. Genet.* **25,** ddw288 (2016).

175. Huuskonen, P. *et al.* Microarray Analysis of the Global Alterations in the Gene Expression in the Placentas From Cigarette-smoking Mothers. *Clin. Pharmacol. Ther.* **83,** 542–550 (2008).

176. Hastie, C., Haw, S. & Pell, J. Impact of smoking cessation and lifetime exposure on C-reactive protein. *Nicotine Tob. Res.* **10,** 637–642 (2008).

177. Fröhlich, M. *et al.* Independent association of various smoking characteristics with markers of systemic inflammation in men. Results from a representative sample of the general population (MONICA Augsburg Survey 1994/95). *Eur. Heart J.* **24,** 1365–72 (2003).

178. Wannamethee, S. G. *et al.* Associations between cigarette smoking, pipe/cigar smoking, and smoking cessation, and haemostatic and inflammatory markers for cardiovascular disease. *Eur. Heart J.* **26,** 1765–1773 (2005).

179. Yasue, H. *et al.* Low-grade inflammation, thrombogenicity, and atherogenic lipid profile in cigarette smokers. *Circ. J.* **70,** 8–13 (2006).

180. Azar, R. & Richard, A. Elevated salivary C-reactive protein

levels are associated with active and passive smoking in healthy youth: A pilot study. *J. Inflamm.* **8,** 37 (2011).

181. Wang, D. *et al.* Exposure to Parental Smoking in Childhood is Associated with High C-Reactive Protein in Adulthood: The Cardiovascular Risk in Young Finns Study. *J. Atheroscler. Thromb.* **24,** 1231–1241 (2017).
182. Airoldi, L. *et al.* Effects of cigarette smoking on the human urinary proteome. *Biochem. Biophys. Res. Commun.* **381,** 397–402 (2009).
183. Yang, M. *et al.* Long-term smoking alters abundance of over half of the proteome in bronchoalveolar lavage cell in smokers with normal spirometry, with effects on molecular pathways associated with COPD. *Respir. Res.* **19,** 40 (2018).
184. Chelchowska, M. *et al.* The pregnancy-associated plasma protein A and insulin-like growth factor system in response to cigarette smoking. *J. Matern. Neonatal Med.* **25,** 2377–2380 (2012).
185. van Otterdijk, S. D., Binder, A. M. & Michels, K. B. Locus-specific DNA methylation in the placenta is associated with levels of pro-inflammatory proteins in cord blood and they are both independently affected by maternal smoking during pregnancy. *Epigenetics* **12,** 875–885 (2017).
186. Huuskonen, P. *et al.* The human placental proteome is affected by maternal smoking. *Reprod. Toxicol.* **63,** 22–31 (2016).
187. Filis, P. *et al.* Maternal Smoking Dysregulates Protein Expression in Second Trimester Human Fetal Livers in a Sex-Specific Manner. *J. Clin. Endocrinol. Metab.* **100,** E861–E870

(2015).

188. Benowitz, N. L., Hukkanen, J., Jacob, P. & III. Nicotine chemistry, metabolism, kinetics and biomarkers. *Handb. Exp. Pharmacol.* 29–60 (2009). doi:10.1007/978-3-540-69248-5_2

189. Matsumoto, A. *et al.* Simultaneous measurement of urinary total nicotine and cotinine as biomarkers of active and passive smoking among Japanese individuals. *Environ. Health Prev. Med.* **18,** 244–50 (2013).

190. Stragierowicz, J., Mikołajewska, K., Zawadzka-Stolarz, M., Polańska, K. & Ligocka, D. Estimation of cutoff values of cotinine in urine and saliva for pregnant women in Poland. *BioMed Research International* **2013,** (2013).

191. Hukkanen, J., Jacob, P. & Benowitz, N. L. Metabolism and disposition kinetics of nicotine. *Pharmacol. Rev.* **57,** 79–115 (2005).

192. Kim, S. Overview of Cotinine Cutoff Values for Smoking Status Classification. *Int. J. Environ. Res. Public Health* doi:10.3390/ijerph13121236

193. Polanska, K. *et al.* Estimation of Saliva Cotinine Cut-Off Points for Active and Passive Smoking during Pregnancy-Polish Mother and Child Cohort (REPRO_PL). *Int. J. Environ. Res. Public Health* **13,** (2016).

194. Mueller, D. C., Piller, M., Niessner, R., Scherer, M. & Scherer, G. Untargeted Metabolomic Profiling in Saliva of Smokers and Nonsmokers by a Validated GC-TOF-MS Method. *J. Proteome Res.* **13,** 1602–1613 (2014).

195. Yuan, J.-M. *et al.* Urinary Levels of Cigarette Smoke

Constituent Metabolites Are Prospectively Associated with Lung Cancer Development in Smokers. *Cancer Res.* **71,** 6749–6757 (2011).

196. Thomas, J. L. *et al.* Metabolites of a Tobacco-Specific Lung Carcinogen in Children Exposed to Secondhand or Thirdhand Tobacco Smoke in Their Homes. *Cancer Epidemiol. Biomarkers Prev.* **20,** 1213–1221 (2011).
197. Anderson, K. E. *et al.* Metabolites of a tobacco-specific lung carcinogen in nonsmoking women exposed to environmental tobacco smoke. *J. Natl. Cancer Inst.* **93,** 378–81 (2001).
198. Hsu, P.-C. *et al.* Feasibility of Identifying the Tobacco-related Global Metabolome in Blood by UPLC–QTOF-MS. *J. Proteome Res.* **12,** 679–691 (2013).
199. Huang, Y. *et al.* Untargeted metabolomics reveals multiple metabolites influencing smoking-related DNA methylation. *Epigenomics* **10,** 379–393 (2018).
200. Xu, T. *et al.* Effects of smoking and smoking cessation on human serum metabolite profile: results from the KORA cohort study. *BMC Med.* **11,** 60 (2013).
201. Kris-Etherton, P. M. AHA Science Advisory. Monounsaturated fatty acids and risk of cardiovascular disease. American Heart Association. Nutrition Committee. *Circulation* **100,** 1253–8 (1999).
202. Mozaffarian, D., Rimm, E. B. & Herrington, D. M. Dietary fats, carbohydrate, and progression of coronary atherosclerosis in postmenopausal women. *Am. J. Clin. Nutr.* **80,** 1175–1184 (2004).

203. Shah, S. H. *et al.* Association of a Peripheral Blood Metabolic Profile With Coronary Artery Disease and Risk of Subsequent Cardiovascular Events. *Circ. Cardiovasc. Genet.* **3,** 207–214 (2010).

204. Asking Questions about Alcohol in Pregnancy (AQUA): Longitudinal cohort study of the effects of low and moderate doses of alcohol exposure on the fetus | FASD Hub. Available at: https://www.fasdhub.org.au/research-and-publications/research-projects/AskingQuestions-alcohol-use-pregnancy/. (Accessed: 5th January 2018)

205. Heude, B. *et al.* Cohort Profile: The EDEN mother-child cohort on the prenatal and early postnatal determinants of child health and development. *Int. J. Epidemiol.* **45,** 353–363 (2016).

206. Guillemette, L. *et al.* Genetics of Glucose regulation in Gestation and Growth (Gen3G): a prospective prebirth cohort of mother–child pairs in Sherbrooke, Canada. *BMJ Open* **6,** e010031 (2016).

207. Birthcohort | Birthcohorts. Available at: http://www.birthcohorts.net/birthcohorts/birthcohort/?id=207. (Accessed: 28th July 2017)

208. Guxens, M. *et al.* Cohort Profile: The INMA—INfancia y Medio Ambiente—(Environment and Childhood) Project. *Int. J. Epidemiol.* **41,** 930–940 (2012).

209. Children's Environmental Health and Disease Prevention Research Center at Dartmouth. Available at: http://www.dartmouth.edu/~childrenshealth/NHBCS.php.

(Accessed: 28th July 2017)

210. STUDY of child health services in Rhode Island. *R. I. Med. J.* **32,** 494–520 (1949).

211. Fortin, J.-P. *et al.* Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biol.* **15,** 503 (2014).

212. Teschendorff, A. E. *et al.* A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics* **29,** 189–96 (2013).

213. Chen, J. *et al.* CpGFilter: model-based CpG probe filtering with replicates for epigenome-wide association studies. *Bioinformatics* **32,** 469–471 (2016).

214. Houseman, E. A., Molitor, J. & Marsit, C. J. Reference-free cell mixture adjustments in analysis of DNA methylation data. *Bioinformatics* **30,** 1431–9 (2014).

215. MASS package | R Documentation. Available at: https://www.rdocumentation.org/packages/MASS/versions/7.3-47. (Accessed: 14th March 2018)

216. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26,** 2190–2191 (2010).

217. Deyssenroth, M. A. *et al.* Whole-transcriptome analysis delineates the human placenta gene network and its associations with fetal growth. *BMC Genomics* **18,** 520 (2017).

218. Bioconductor - MEAL. Available at:

http://bioconductor.org/packages/release/bioc/html/MEAL.html. (Accessed: 17th January 2018)

219. Rahmani, E. *et al.* Genome-wide methylation data mirror ancestry information. *Epigenetics Chromatin* **10,** 1 (2017).

220. Bioconductor - IlluminaHumanMethylation450kanno.ilmn12.hg19. Available at: https://bioconductor.org/packages/release/data/annotation/html/IlluminaHumanMethylation450kanno.ilmn12.hg19.html. (Accessed: 16th January 2018)

221. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* **9,** 215–6 (2012).

222. Hamada, H. *et al.* Allele-Specific Methylome and Transcriptome Analysis Reveals Widespread Imprinting in the Human Placenta. *Am. J. Hum. Genet.* **99,** 1045–1058 (2016).

223. ConsensusPathDB.

224. Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool.

225. Horikoshi, M. *et al.* Genome-wide associations for birth weight and correlations with adult disease. *Nature* **538,** 248–252 (2016).

226. Beaumont, R. N. *et al.* Genome-wide association study of offspring birth weight in 86 577 women identifies five novel loci and highlights maternal genetic effects that are independent of fetal genetics. *Hum. Mol. Genet.* **27,** 742–756 (2018).

227. van der Valk, R. J. P. *et al.* A novel common variant in DCST2 is associated with length in early life and height in adulthood. *Hum. Mol. Genet.* **24,** 1155–68 (2015).
228. Rob Taal, H. *et al.* Common variants at 12q15 and 12q24 are associated with infant head circumference. *Mònica Guxens Marjan Kerkhof Nicholas J Timpson Lawrence J Beilin Klaus Bønnelykke Jessica L Buxt. Pimphen Charoen Johan Eriksson Xavier Estiv. Mads Melbye Hans Bisgaard Craig E Pennell Elisabeth Widen* doi:10.1038/ng.2238
229. Zhang, G. *et al.* Genetic Associations with Gestational Duration and Spontaneous Preterm Birth. *N. Engl. J. Med.* **377,** 1156–1167 (2017).
230. Package 'MASS' Title Support Functions and Datasets for Venables and Ripley's MASS. (2018).
231. Villar, J. *et al.* International standards for newborn weight, length, and head circumference by gestational age and sex: the Newborn Cross-Sectional Study of the INTERGROWTH-21st Project. *Lancet* **384,** 857–868 (2014).
232. Muggli, E. *et al.* Study protocol: Asking QUestions about Alcohol in pregnancy (AQUA): a longitudinal cohort study of fetal effects of low to moderate alcohol exposure. *BMC Pregnancy Childbirth* **14,** 302 (2014).
233. Wright, J. *et al.* Cohort profile: The born in bradford multi-ethnic family cohort study. *Int. J. Epidemiol.* (2013). doi:10.1093/ije/dys112
234. Heude, B. *et al.* Cohort Profile: The EDEN mother-child cohort on the prenatal and early postnatal determinants of child

health and development. *Int. J. Epidemiol.* **45,** 353–363 (2016).

235. Guxens, M. *et al.* Cohort Profile: The INMA--INfancia y Medio Ambiente--(Environment and Childhood) Project. *Int. J. Epidemiol.* **054,** (2011).
236. Grazuleviciene, R., Danileviciute, A., Nadisauskiene, R. & Vencloviene, J. Maternal smoking, GSTM1 and GSTT1 polymorphism and susceptibility to adverse pregnancy outcomes. *Int. J. Environ. Res. Public Health* **6,** 1282–97 (2009).
237. Magnus, P. *et al.* Cohort Profile Update: The Norwegian Mother and Child Cohort Study (MoBa). *Int. J. Epidemiol.* **45,** 382–388 (2016).
238. Chatzi, L. *et al.* Cohort Profile: The Mother-Child Cohort in Crete, Greece (Rhea Study). *Int. J. Epidemiol.* **46,** 1392–1393k (2017).
239. ESCAPE Project: long-term effects on human health of exposure to air pollution in Europe. Available at: http://www.escapeproject.eu/. (Accessed: 27th May 2018)
240. Sandoval, J. *et al.* Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics* **6,** 692–702 (2011).
241. van Iterson, M. *et al.* MethylAid: visual and interactive quality control of large Illumina 450k datasets. *Bioinformatics* **30,** 3435–7 (2014).
242. Lehne, B. *et al.* A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data

quality and performance in epigenome-wide association studies. *Genome Biol.* **16,** 37 (2015).

243. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8,** 118–127 (2007).
244. Chen, Y. *et al.* Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics* **8,** 203–9 (2013).
245. MagMAX ™ for Stabilized Blood Tubes RNA Isolation Kit (Compatible with Tempus ® Blood RNA Tubes) Protocol For Research Use Only. Not intended for any animal or human therapeutic or diagnostic use. (2010).
246. Affymetrix & Inc. *User Manual GeneChip ® Whole Transcript (WT) Sense Target Labeling Assay*.
247. Affymetrix. Available at: http://www.affymetrix.com/support/help/exon_glossary/index.affx. (Accessed: 27th August 2018)
248. Package 'isva' Type Package Title Independent Surrogate Variable Analysis. (2017). doi:10.1093/bioinformatics/btr171
249. *Package 'SmartSVA'*. (2017).
250. Universal microRNA Reference Kit from Agilent Technologies | Biocompare.com. Available at: https://www.biocompare.com/23751-Universal-Mouse-Reference-RNA/1833198-Universal-microRNA-Reference-Kit/. (Accessed: 23rd June 2018)
251. Kozomara, A. & Griffiths-Jones, S. miRBase: annotating high

confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* **42,** D68–D73 (2014).

252. Suo, C., Salim, A., Chia, K.-S., Pawitan, Y. & Calza, S. Modified least-variant set normalization for miRNA microarray. *RNA* **16,** 2293–303 (2010).
253. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43,** e47–e47 (2015).
254. Bioconductor - limma. Available at: https://bioconductor.org/packages/release/bioc/html/limma.html. (Accessed: 10th August 2018)
255. Bioconductor - sva. Available at: https://bioconductor.org/packages/release/bioc/html/sva.html. (Accessed: 27th August 2018)
256. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28,** 882–883 (2012).
257. Nadarajah, S. & Kotz, S. The Exponentiated Type Distributions. *Acta Appl. Math.* **92,** 97–111 (2006).
258. Langsdorf, M. *et al.* Metabolomic Analysis using the AbsoluteIDQ™ p150 Kit in combination with the AB Sciex 5500 QTrap ® Mass Spectrometer.
259. Veselkov, K. A. *et al.* Recursive segment-wise peak alignment of biological (1)h NMR spectra for improved metabolic biomarker recovery. *Anal. Chem.* **81,** 56–66 (2009).
260. Tredwell, G. D., Behrends, V., Geier, F. M., Liebeke, M. &

Bundy, J. G. Between-Person Comparison of Metabolite Fitting for NMR-Based Quantitative Metabolomics. *Anal. Chem.* **83,** 8683–8687 (2011).

261. Cloarec, O. *et al.* Statistical Total Correlation Spectroscopy: An Exploratory Approach for Latent Biomarker Identification from Metabolic ^{1}H NMR Data Sets. *Anal. Chem.* **77,** 1282–1289 (2005).
262. Li, M.-X., Yeung, J. M. Y., Cherny, S. S. & Sham, P. C. Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Hum. Genet.* **131,** 747–56 (2012).
263. R: Tetrachoric, polychoric, biserial and polyserial correlations... Available at: https://www.personality-project.org/r/psych/help/tetrachor.html. (Accessed: 16th September 2018)
264. Textor, J., van der Zander, B., Gilthorpe, M. S., Liśkiewicz, M. & Ellison, G. T. H. Robust causal inference using directed acyclic graphs: the R package 'dagitty'. *Int. J. Epidemiol.* dyw341 (2017). doi:10.1093/ije/dyw341
265. Body mass index - BMI. (2018).
266. WHO | BMI-for-age (5-19 years). *WHO* (2015).
267. de Onis, M. *et al.* Development of a WHO growth reference for school-aged children and adolescents. *Bull. World Health Organ.* **85,** 660–7 (2007).
268. Houseman, E. A. *et al.* Reference-free deconvolution of DNA methylation data and mediation by cell composition effects.

BMC Bioinformatics **17,** 259 (2016).

269. Reinius, L. E. *et al.* Differential DNA Methylation in Purified Human Blood Cells: Implications for Cell Lineage and Studies on Disease Susceptibility. *PLoS One* **7,** e41361 (2012).
270. White, I. R., Royston, P. & Wood, A. M. Multiple imputation using chained equations: Issues and guidance for practice. *Stat. Med.* **30,** 377–399 (2011).
271. Buuren, S. van & Groothuis-Oudshoorn, K. **mice** : Multivariate Imputation by Chained Equations in *R*. *J. Stat. Softw.* **45,** 1–67 (2011).
272. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57,** 289–300 (1995).
273. Bioconductor - omicRexposome. Available at: https://bioconductor.org/packages/release/bioc/html/omicRexposome.html. (Accessed: 10th August 2018)
274. Hernandez-Ferrer, C., Ruiz-Arenas, C., Beltran-Gomila, A. & González, J. R. MultiDataSet: an R package for encapsulating multiple data sets with application to omic data integration. *BMC Bioinformatics* **18,** 36 (2017).
275. Bioconductor - rexposome. Available at: https://bioconductor.org/packages/release/bioc/html/rexposome.html. (Accessed: 10th August 2018)
276. *Package 'ggplot2' Title Create Elegant Data Visualisations Using the Grammar of Graphics*. (2018).
277. Stephen Turner, A. & Stephen Turner, M. *Package 'qqman'*

Title Q-Q and Manhattan Plots for GWAS Data. (2017).

278. Jan Graffelman, A. & Jan Graffelman, M. Package 'calibrate' Title Calibration of Scatterplot and Biplot Axes. (2015).
279. Mook-Kanamori, D. O. *et al.* Variants near CCNL1/LEKR1 and in ADCY5 and fetal growth characteristics in different trimesters. *J. Clin. Endocrinol. Metab.* **96,** E810-5 (2011).
280. Urbanek, M. *et al.* The chromosome 3q25 genomic region is associated with measures of adiposity in newborns in a multi-ethnic genome-wide association study. *Hum. Mol. Genet.* **22,** 3583–96 (2013).
281. Hivert, M.-F. *et al.* Genetic determinants of adiponectin regulation revealed by pregnancy. *Obesity* **25,** 935–944 (2017).
282. Andersson, E. A. *et al.* The Birth Weight Lowering C-Allele of rs900400 Near LEKR1 and CCNL1 Associates with Elevated Insulin Release following an Oral Glucose Challenge. *PLoS One* **6,** e27096 (2011).
283. Horikoshi, M. *et al.* Genome-wide associations for birth weight and correlations with adult disease. *Nature* **538,** 248–252 (2016).
284. Li, C. *et al.* Genome-Wide Association Study Meta-Analysis of Long-Term Average Blood Pressure in East Asians. *Circ. Cardiovasc. Genet.* **10,** e001527 (2017).
285. Daoud, G., Le bellego, F. & Lafond, J. PP2 regulates human trophoblast cells differentiation by activating p38 and ERK1/2 and inhibiting FAK activation. *Placenta* **29,** 862–70 (2008).
286. Chan, M. C. *et al.* Tuning the Transcriptional Response to

Hypoxia by Inhibiting Hypoxia-inducible Factor (HIF) Prolyl and Asparaginyl Hydroxylases. *J. Biol. Chem.* **291,** 20661–20673 (2016).

287. Chen, D. Regulation of Placental Angiogenesis. (2013).
288. Sahay, A. S., Sundrani, D. P. & Joshi, S. R. Neurotrophins: Role in Placental Growth and Development. *Vitam. Horm.* **104,** 243–261 (2017).
289. Jauniaux, E. & Burton, G. J. Morphological and biological effects of maternal exposure to tobacco smoke on the feto-placental unit. *Early Hum. Dev.* **83,** 699–706 (2007).
290. Zdravkovic, T., Genbacev, O., McMaster, M. T. & Fisher, S. J. The adverse effects of maternal smoking on the human placenta: A review. *Placenta* **26,** S81–S86 (2005).
291. Paul, S., Home, P., Bhattacharya, B. & Ray, S. GATA factors: Master regulators of gene expression in trophoblast progenitors. *Placenta* **60,** S61–S66 (2017).
292. Gonçalves, R. B. *et al.* Impact of smoking on inflammation: Overview of molecular mechanisms. *Inflamm. Res.* **60,** 409–424 (2011).
293. Perez-Garcia, V. *et al.* Placentation defects are highly prevalent in embryonic lethal mouse mutants. *Nature* **555,** 463–468 (2018).
294. Dior, U. P. *et al.* Parental smoking during pregnancy and offspring cardio-metabolic risk factors at ages 17 and 32. *Atherosclerosis* **235,** 430–7 (2014).
295. Behl, M. *et al.* Evaluation of the Association between Maternal Smoking, Childhood Obesity, and Metabolic

Disorders: A National Toxicology Program Workshop Review. *Environ. Health Perspect.* **121,** 170–180 (2012).

296. Beaumont, R. N. *et al.* Genome-wide association study of offspring birth weight in 86 577 women identifies five novel loci and highlights maternal genetic effects that are independent of fetal genetics. *Hum. Mol. Genet.* **27,** 742–756 (2018).

297. Gao, X., Thomsen, H., Zhang, Y., Breitling, L. P. & Brenner, H. The impact of methylation quantitative trait loci (mQTLs) on active smoking-related DNA methylation changes. *Clin. Epigenetics* **9,** 87 (2017).

298. Dogan, M. V., Beach, S. R. H. & Philibert, R. A. Genetically contextual effects of smoking on genome wide DNA methylation. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* **174,** 595–607 (2017).

299. Joubert, B. R. *et al.* DNA Methylation in Newborns and Maternal Smoking in Pregnancy: Genome-wide Consortium Meta-analysis. *Am. J. Hum. Genet.* **98,** 680–696 (2016).

300. Tsai, P.-C. *et al.* Smoking induces coordinated DNA methylation and gene expression changes in adipose tissue with consequences for metabolic health. *bioRxiv* 353581 (2018). doi:10.1101/353581

301. Gao, X., Jia, M., Zhang, Y., Philipp Breitling, L. & Brenner, H. DNA methylation changes of whole blood cells in response to active smoking exposure in adults: a systematic review of DNA methylation studies. (2011). doi:10.1186/s13148-015-0148-3

302. Neidhart, M. *DNA methylation and complex human disease*.
303. Küpers, L. K. *et al.* DNA methylation mediates the effect of maternal smoking during pregnancy on birthweight of the offspring. *Int. J. Epidemiol.* **44,** 1224–1237 (2015).
304. Joubert, B. R. *et al.* DNA Methylation in Newborns and Maternal Smoking in Pregnancy: Genome-wide Consortium Meta-analysis. *Am. J. Hum. Genet.* **98,** 680–96 (2016).
305. CNTNAP2 gene - Genetics Home Reference - NIH. Available at: https://ghr.nlm.nih.gov/gene/CNTNAP2. (Accessed: 18th September 2018)
306. Rzehak, P. *et al.* Maternal Smoking during Pregnancy and DNA-Methylation in Children at Age 5.5 Years: Epigenome-Wide-Analysis in the European Childhood Obesity Project (CHOP)-Study. *PLoS One* **11,** e0155554 (2016).
307. TDGF1 Gene - GeneCards | TDGF1 Protein | TDGF1 Antibody. Available at: https://genecards.weizmann.ac.il/v3/cgi-bin/carddisp.pl?gene=TDGF1. (Accessed: 31st August 2018)
308. Zhao, Y., Fu, D., Xu, C., Yang, J. & Wang, Z. Identification of genes associated with tongue cancer in patients with a history of tobacco and/or alcohol use. *Oncol. Lett.* **13,** 629–638 (2017).
309. Simpson, A. J., Gray, R. S., Moore, N. R. & Booth, N. A. The effects of chronic smoking on the fibrinolytic potential of plasma and platelets. *Br. J. Haematol.* **97,** 208–13 (1997).
310. Sasaki, A., Kurisu, A., Ohno, M. & Ikeda, Y. Overweight/obesity, smoking, and heavy alcohol consumption

are important determinants of plasma PAI-1 levels in healthy men. *Am. J. Med. Sci.* **322,** 19–23 (2001).

311. Ozaki, K., Hori, T., Ishibashi, T., Nishio, M. & Aizawa, Y. Effects of chronic cigarette smoking on endothelial function in young men. *J. Cardiol.* **56,** 307–313 (2010).
312. GeneCards - Human Genes | Gene Database | Gene Search. Available at: https://www.genecards.org/. (Accessed: 7th August 2018)
313. Ma, Q. *et al.* Genetic variants in PLG, LPA, and SIGLEC 14 as well as smoking contribute to plasma plasminogen levels. *Blood* **124,** 3155–64 (2014).
314. Hanna, C. W. *et al.* Pervasive polymorphic imprinted methylation in the human placenta. *Genome Res.* **26,** 756–67 (2016).
315. Fa, S. *et al.* Changes in first trimester fetal CYP1A1 and AHRR DNA methylation and mRNA expression in response to exposure to maternal cigarette smoking. *Environ. Toxicol. Pharmacol.* **57,** 19–27 (2018).
316. Wang, Y. & Zhao, S. Cell Types of the Placenta. (2010).
317. Fogarty, N. M. E., Burton, G. J. & Ferguson-Smith, A. C. Different epigenetic states define syncytiotrophoblast and cytotrophoblast nuclei in the trophoblast of the human placenta. *Placenta* **36,** 796–802 (2015).
318. Bauer, M. *et al.* Tobacco smoking differently influences cell types of the innate and adaptive immune system—indications from CpG site methylation. *Clin. Epigenetics* **8,** 83 (2016).
319. Davey Smith, G. & Hemani, G. Mendelian randomization:

genetic anchors for causal inference in epidemiological studies. *Hum. Mol. Genet.* **23,** R89-98 (2014).