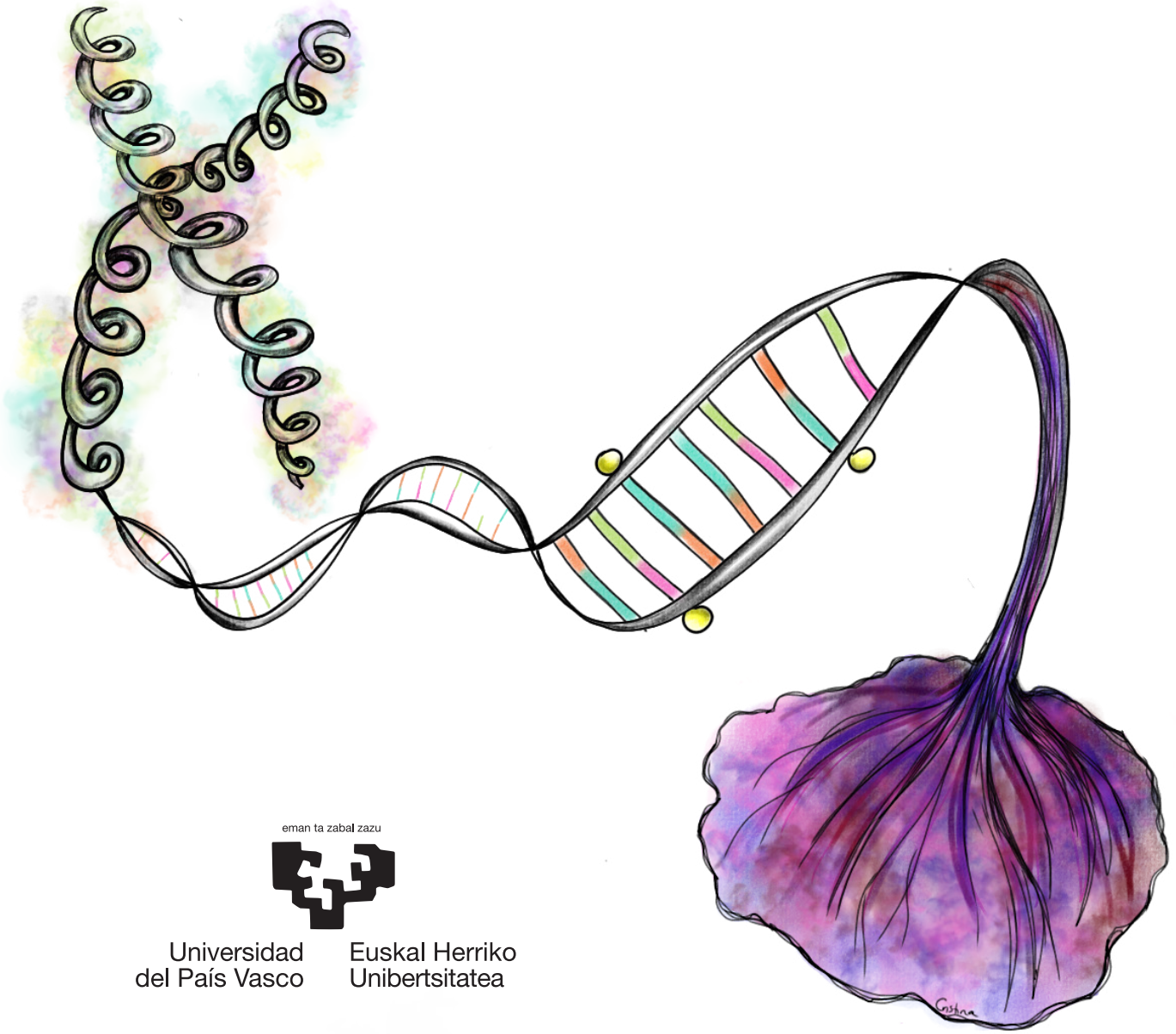
Prenatal crosstalk between environment and genetics: the impact of placental DNA methylation on future health

Ariadna Cilleros Portet

Leioa, 2024



eman ta zabal zazu

Universidad del País Vasco

Euskal Herriko Unibertsitatea

eman ta zabal zazu

Universidad del País Vasco

Euskal Herriko Unibertsitatea

Prenatal crosstalk between environment and genetics: the impact of placental DNA methylation on future health

PhD Thesis

Ariadna Cilleros Portet

Supervisors

Nora Fernández Jiménez & Jose Ramon Bilbao Catalá

This thesis has been submitted to the doctoral program in Molecular Biology and Biomedicine at the University of the Basque Country.

Leioa, 2024

Acknowledgements

I would like to thank my supervisors, Nora and Jose Ramon, for their dedication and supervision throughout this thesis. Thank you for trusting me from the very beginning when I had just finished my degree. Especially, thank you Nora for allowing me to participate in your project on placental mQTLs, seeing me as an equal by always considering my opinion, and finally for taking care of me as if I was part of your family. Thank you both of you and the whole team for supporting this work. Eskerrik asko!

I would also like to thank the other researchers from the INMA project, such as Loreto Santa Marina, Jesús Ibarluzea and Mariona Bustamante, for the huge work with the participants, the sample requirement, data acquisition, and helpful discussions. As well as Corina Lesseur and Johanna Lepeule, thank you for welcoming me into your research groups. I cannot overlook people like Malu Calle, who has had a great impact on my professional life, and with whom I hope to continue enjoying many more coffees full of science and appreciation.

Mi tesis doctoral no habría llegado a este punto sin personas como las que menciono a continuación. Por un lado, gracias de todo corazón a todas las componentes de Reperkusion Feminista; Aiala, Ainhoa, Anna, Andrea, Anita, Belén, Esther, Gemma, Iria, Jessi, Joana, June, Lorea, Lutxi, Maider, Maite B.U., Maite B.Z., Marta A., Marta F., Mila, Nekane, Nerea, Nuria F.L., Núria F.G. e Yvonne. Cada una de vosotras forma parte de mí y de esta tesis, gracias por cada abrazo, por cada carcajada y por cada momento vivido a vuestro lado. Sois el hogar que nunca abandonaré. Muchas gracias, eskerrik asko, moltes gràcies.

La otra mitad de mi corazón bilbaíno pertenece a la familia Vink-Larruskain, que desde hace cuatro años me acogieron en su casa como una hija y una hermana más. Mi tercer y cuarto apellidos son Vink-Larruskain, mi *ama* se llama Maite, mi *aita* se llama Víctor, y mis hermanos son Naiara e Íñigo. Gracias Maite, Víctor, Naiara, Íñigo, Mauri, Gorka, Olivia y Marco, las comidas de los domingos siempre serán en San Ignazio acompañados con las imprescindibles rabas del Presley. Os quiero.

També m'agradaria tenir un petit detall des d'aquí a una peça fonamental de la meva salut, la meva psicòloga, l'Olga. Si ve la meva dedicació a la salut mental i neuropsiquiàtrica al llarg d'aquest projecte no és atzarosa, és perquè des de fa molts anys que en soc conscient de la seva importància. L'Olga m'ha ajudat a superar aquesta etapa dotant-me d'eines imprescindibles, no només pel doctorat, sinó per tot el transcurs de la meva vida, tant en l'àmbit professional, com en el personal. Des d'aquí, moltes gràcies.

Agradecer desde aquí a Cristina Iribarren por su preciosa ilustración y amistad. Gracias Cristina por haber estado en uno de los momentos más difíciles de mi vida personal, gracias por tu escucha paciente, los cafés y los fines de semana paseando por las calles de Nueva York. Este doctorado también me ha permitido conocer a personas maravillosas como Arantza, Anna, Bárbara o Rianne. ¡A todas vosotras, gracias y viva la mujer en la ciencia!

No vull passar per alt el suport de totes les persones que han set al meu costat al llarg d'aquest projecte: a la família materna i paterna, així com l'Alba, en Marc o l'Uxue. Moltes gràcies!

Finalment, m'agradaria dedicar aquesta tesi al meu germà, Gerard, el meu pare, Miquel, i en especial a la meva mare, Esther. Gràcies Gerard per ser un exemple de valentia, no deixis mai d'apuntar alt, però si et despistes em trobaràs ballant i molestant-te a l'entrada de la teva habitació. Segueix perseguint les idees que tant anheles independentment d'on et portin, nosaltres seguirem al teu costat. Gràcies papa pel suport incondicional, l'amor inescrutable que em demostres dia a dia, les roses i els llibres de Sant Jordi a distància, i tots els viatges de Tona a l'aeroport, aquells moments en què de cop el cotxe es convertia en un lloc segur on confiar totes les nostres inquietuds. I per acabar, gràcies mama, a tu també agrair-te el suport incondicional i l'amor inescrutable. De tu he après què era l'esforç i la perseverança. Gràcies per haver set una referent a seguir, perquè tal com em vas dir tu un dia *la gota d'aigua no forada la roca per la seva força, sinó per la seva constància.* Us estimo.

[...] Perquè cap ho va fer sol. Sis dones van quedar fora de camp de la foto.
Tres també serien arquitectes. Dues, interioristes. Una, pintora.
I totes amb tanta o més fal·lera per les cases de pagès que ells.
I amb tanta o més destresa per rescabalar-les que ells.
Quan es grata sempre se les troba.
Però s'ha de gratar

Les nostres mares, de Gemma Ruiz Palà

The aim of this work is to elucidate the molecular mechanisms by which health and disease are developed. However, the aim has never been to increase the social burden imposed to mothers and women. Studies in many other factors other than pregnancy and maternity should be carried out in order to improve the health of both mothers and children.

Index

Main publications related to this thesis

Fernandez-Jimenez, N.*, Fore, R.*, **Cilleros-Portet, A.*** et al. (**Equally contributed*). A meta-analysis of pre-pregnancy maternal body mass index and placental DNA methylation identifies 27 CpG sites with implications for mother-child health. *Communications Biology* **5**, 1313 (2022).

Cilleros-Portet, A. et al. Potentially causal associations between placental DNA methylation and schizophrenia and other neuropsychiatric disorders. *medRxiv (Under review in Nature Communications)* (2023).

Other publications

González-García, B. P., Marí, S., **Cilleros-Portet, A.**, et al. Two-Sample Mendelian Randomization detects bidirectional causality between gut microbiota and celiac disease in individuals with high genetic risk. *Frontiers in Immunology* **14**, 1082862 (2023).

Hernangomez-Laderas, A., **Cilleros-Portet, A.**, et al. Sex bias in celiac disease: XWAS and monocyte eQTLs in women identify *TMEM187* as a functional candidate gene. *Biology of Sex Differences* **14**, 86 (2023).

Cosin-Tomas, M., **Cilleros-Portet, A.** et al. Prenatal Maternal Smoke, DNA Methylation, and Multi-omics of Tissues and Child Health. *Current Environmental Health Reports* **9**, 502-512 (2022).

Rueda-Martínez, A., Garitazelaia, A., **Cilleros-Portet, A.** et al. Genetic Contribution of Endometriosis to the Risk of Developing Hormone-Related Cancers. *International Journal of Molecular Sciences* **22**, 6083 (2021).

Garitazelaia, A., Rueda-Martínez, A., Arauzo, R., de Miguel, J., **Cilleros-Portet, A.** et al. A Systematic Two-Sample Mendelian Randomization Analysis Identifies Shared Genetic Origin of Endometriosis and Associated Phenotypes. *Life (Basel)* **11**, 24 (2021).

García-Santisteban, I., Romero-Garmendia, I., **Cilleros-Portet, A.**, et al. Celiac disease susceptibility: The genome and beyond. *International Review of Cell and Molecular Biology* **358**, 1-45 (2021).

García-Santisteban, I., **Cilleros-Portet, A.** et al. A Two-Sample Mendelian Randomization Analysis Investigates Associations Between Gut Microbiota and Celiac Disease. *Nutrients* **12**,1420 (2020).

Abbreviations

1000G	1000 Genomes project
ADHD	attention-deficit and hyperactivity disorder
AGR	aggression
AQUA	Asking Questions about Alcohol in pregnancy study
ALSPAC	Avon Longitudinal Study of Parents and Children
ASD	autism spectrum disorder
BIP	bipolar disorder
BMI	body mass index
BMIQ	beta-mixture quantile normalization
BW	birthweight
caQTL	chromatin accessibility quantitative trait locus
CpG	cytosine-phosphate-guanine dinucleotide
DHS	DNase I hypersensitivity site
DMP	differentially methylated position
DNAm	DNA methylation
DO	Disease Ontology
DOHaD	developmental origins of health and disease
EAGLE	Early Genetics and Lifecourse Epidemiology study
EARLI	Early Autism Risk Longitudinal Investigation study
EDEN	study on the pre- and early postnatal determinants of child health and development *(Étude de cohorte généraliste menée en France sur les Déterminants pré- et postnatals précoces du développement psychomoteur et de la santé de l'ENfant*)
eFORGE	experimentally-derived Functional element Overlap analysis of ReGions from EWAS tool
ELC	early life complications
ENCODE	Encyclopedia of DNA Elements project
eQTL	expression quantitative trait locus
eQTM	expression quantitative trait methylation
EWAS	epigenome-wide association study
FDR	False Discovery Rate

GA	gestational age
GCTA	Genome-wide Complex Trait Analysis
Gen3G	Genetics of Glucose regulation in Gestation and Growth study
GENEIDA	Genetics, Early Life Environmental Exposures and Infant Development in Andalusia study
GO	Gene Ontology
GTEx	Genotype-Tissue Expression project
GWAMA	Genome-Wide Association Meta Analysis
GWAS	genome-wide association study
HCMV	human cytomegalovirus
HEBC	Harvard Epigenetic Birth Cohort study
HEIDI	heterogeneity in dependent instruments
HIV	human immunodeficiency virus
HLA	human leukocyte antigen
HPA	hypothalamic-pituitary-adrenal axis
HRC	Haplotype Reference Consortium
HWE	Hardy-Weinberg Equilibrium
ieQTL	interacting expression quantitative trait locus
imQTL	interacting methylation quantitative trait locus
imSite	CpG site participating in the imQTL
imVariant	SNP participating in the imQTL
INMA	Environment and Childhood study (*INfancia y Medio Ambiente*)
INT	internalizing behavior
IPV	intimate partner violence
ITU	InTraUterine sampling in early pregnancy study
IV	instrumental variable
KEGG	Kyoto Encyclopedia of Genes and Genomes
LD	linkage disequilibrium
MAF	minor allele frequency
MARBLES	Markers of Autism Risk in Babies Learning Early Signs study
MDD	major depression disorder
MIA	maternal immune activation
MoBA	Norwegian Mother, Father and Child Cohort study

molQTL	molecular quantitative trait locus
mQTL	methylation quantitative trait locus
MR	mendelian randomization
MSDP	maternal smoking during pregnancy
mSite	CpG site participating in the mQTL
mVariant	SNP participating in the mQTL
N_{eff}	effective number of independent variants in the *cis* window estimated by eigenMT
NHBCS	New Hampshire Birth Cohort Study
nRBC	nucleated red blood cells
OCD	obsessive compulsive disorder
PACE	Pregnancy And Childhood Epigenetics consortium
PC	principal component
PCA	principal component analysis
PD	panic disorder
PGC	Psychiatric Genomics Consortium
ppBMI	pre-pregnancy body mass index
PMD	partially methylated domain
PRS	polygenic risk score
PP	posterior probability
QTL	quantitative trait locus
RICHS	Rhode Island Child Health Study
RNT	rank-based inverse normal transformation
SA	suicidal attempt
SCZ	schizophrenia
SD	standard deviation
STB	syncytiotrophoblast
SMR	summary-based mendelian randomization
SMR-multi	multi-SNP-based MR
SNP	single-nucleotide polymorphism
TB	trophoblast
TF	transcription factor
tRNA	transfer RNA
UCSC	University of California Santa Cruz

INTRODUCTION

1. The Developmental Origins of Health and Disease hypothesis

The Developmental Origins of Health and Disease (DOHaD) hypothesis, initially called Fetal Origins of Adult Disease[1], was coined by Barker in 2007[2]. This hypothesis proposes that perinatal and early life environment can impact fetal and later life health[2]. A wide range of environmental exposures have been studied, including nutrition/metabolic milieu[3], maternal smoking[4], alcohol consumption[5,6], stress[7], infection[8] and socioeconomic determinants such us maternal education[9]. In turn, some studied adverse health outcomes are neurological/cognitive disorders[10], cancer[11], respiratory diseases[12], metabolic diseases[13], inflammatory and immune disorders[14], and cardiovascular conditions[15]. For example, it is well known that prenatal stress affects the quality of the intrauterine environment, and is highly related to cardiovascular and metabolic, as well as behavioral and neurodevelopmental disorders[16].

In this context, it has been proposed that the previously mentioned environmental insults contribute to the uterus environment and consequently impact both fetal development and later life health through interactions with or effects on placenta[17,18].

1.1. The role of placenta in DOHaD

Placenta is the first organ to develop and plays a key role throughout pregnancy coordinating the exchange of nutrients, gases, waste and endocrine signals between mother and fetus[19]. Thus, placenta is an ephemeral organ that is uniquely situated to evaluate prenatal exposures in the context of the DOHaD hypothesis. This organ is comprised of several parts and structures (Figure 1). The maternal and fetal arteries bring oxygenated blood from the mother and fetus, respectively, to the placenta. The maternal and fetal veins collect deoxygenated blood from placenta and return to the

mother and the fetus, respectively. The intervillous space is where the nutrient, gas, waste and endocrine signal exchanges occur. The umbilical cord connects the fetus to the placenta, supplies oxygen and nutrients to the fetus, and removes waste products. Lastly, the chorionic villi increase the surface area for the exchange of materials[20]. The chronic villus is where the villous trees are located. These are the main structures of the placenta in terms of material exchange[21]. These features make the placenta the master regulator of the prenatal milieu. In fact, it constitutes the interface between mother and child during pregnancy. Additionally, in 2022 Bhattacharya and colleagues demonstrated the profound health impact of placental genomic regulation in early life traits, that may persist later in life as etiologic antecedents for complex traits, thus programming development across the life course[22]. Some examples include the genetic regulation of the placental transcriptome underlying birthweight (BW), risk of childhood obesity[23] and fetal neurodevelopment[24,25] as well as disturbances in the development of the placenta on a genetic level, inhibiting fetus growth and wellbeing[26].

Placental omics, including transcriptomics and epigenomics, which are partly regulated by genetics, has been proven useful to study the numerous relationships between prenatal exposures and fetal and early life health outcomes[27].

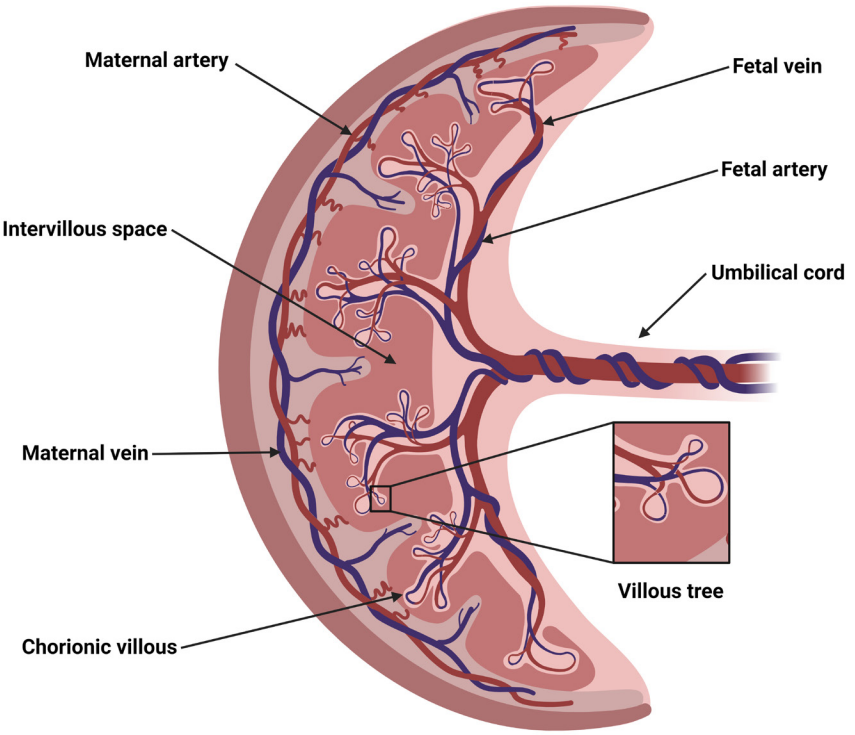


Figure 1. Human placenta structure.

Figure adapted from Mortillo *et al.* 2023[20]. It illustrates the placenta structure highlighting maternal and fetal arteries, maternal and fetal veins, the intervillous space, the umbilical cord, and the chorionic villus where villous trees are located.

2. Epigenetics, the bridge between genetics and environmental factors

The word epigenetics means “in addition to changes in genetic sequence”. The term has evolved to include any process that alters gene activity without changing the DNA sequence and leads to modifications of the genome that can be transmitted to descendant cells[28]. Many types of epigenetic processes have been identified, including histone acetylation and phosphorylation, as well as DNA and histone methylation. Epigenetic processes are natural and essential to many organism functions, but if they occur improperly, major adverse health and behavioral effects can take place. Additionally, these mechanisms mediate the diversified gene expression profiles in the different cells and tissues in multicellular organisms. One of the most characterized epigenetic modifications is DNA methylation (DNAm)[29]. DNAm is an epigenetic mechanism involving the transfer of a methyl group onto position C5 of the cytosine molecule to form 5-methylcytosine (Figure 2)[30]. Most DNAm occurs on cytosines that precede a guanine nucleotide, and form a CpG, or cytosine-phosphate-guanine dinucleotide site, that are spread out across the genome and are heavily methylated except for CpG islands[31]. Briefly, CpG islands are short, dispersed regions of unmethylated DNA with a high frequency of CpG dinucleotides relative to the bulk genome, particularly located in promoter regions[32]. In this context, the terms hypomethylated and hypermethylated are commonly used to refer to an overall decrease and increase, respectively, of the level of 5-methylcystosine in a particular CpG site or region[33]. However, CpG islands are not the only places where DNAm occurs. Methylation has also been observed in shores, shelves, and open sea regions, which are regions that are at a varying distance from a given CpG island[34] (Figure 2). Shores are up to 2 kb from a CpG island, while shelves are from 2 to 4 kb from the island. Open sea refers to isolated regions that do not have a specific designation. This epigenetic modification regulates gene expression by enabling the recruitment of proteins involved in gene regulation or by inhibiting the binding of transcription factors (TF) to DNA[35]. The regulation of gene expression is crucial during

development[36] and in this context, DNAm has been considered a bridge between the environment and the genome, that at the same time, is under the control of both environmental and genetic factors. And since the etiology and pathophysiology of complex traits are incompletely explained individually by any of these two factors, researchers are now exploring the epigenome as the biological interface between the two.

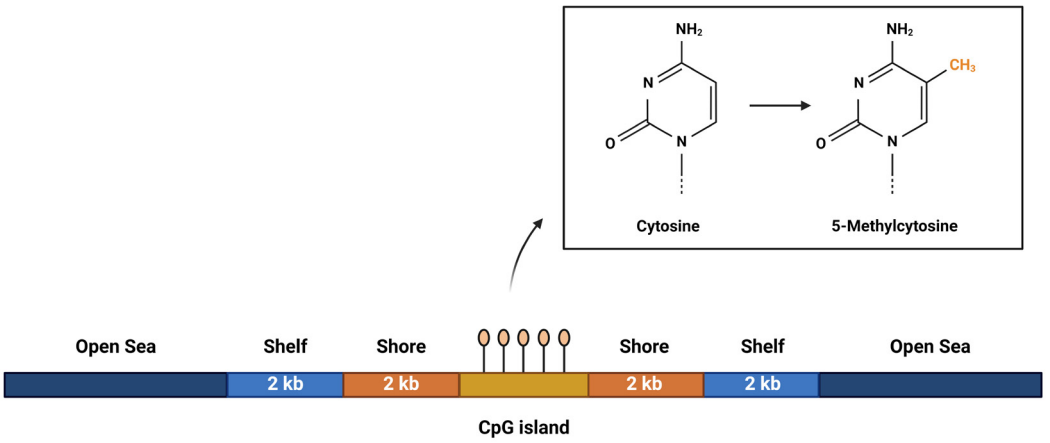


Figure 2. Schematic representation of a CpG island and its surrounding regions.

Non-methylated and methylated CpG site chemistry structures are depicted in the upper panel, the methyl group is highlighted in orange. In the lowest part of the figure the CpG sites are illustrated as a round sign on a pole within the CpG island.

It has been observed that DNAm undergoes extensive alterations not only throughout gestation, but also in response to changing *in utero* conditions[37]. Nowadays there are many studies proposing that the intrauterine environment alters placental function through DNAm[38]. It is important to note that, while in most tissues and cell types DNAm is bimodally distributed with most of the CpG sites hypomethylated or hypermethylated, placental DNAm follows a trimodal distribution, mostly due to its high content of both partially methylated domains (PMD) and CpG positions with intermediate methylation levels[39]. Shortly, PMDs are large (< 100 kb) regions of reduced DNAm interspersed with regions of higher DNAm. Placental genes within

PMDs tend to be tissue-specific and show higher promoter DNAm and reduced expression as compared with somatic tissues[40]. For example, in 2016 a prospective study in autism spectrum disorder (ASD) siblings reported an association between pesticides professionally applied outside home during pregnancy and a higher average methylation over PMDs[41]. However, the particular characteristics of placental DNAm are not only observed in bulk tissue, but also in placental cell types. In 2021, Yuan and colleagues characterized cell composition and cell-specific DNAm dynamics across gestation considering six cell types, namely syncytiotrophoblasts (STB), trophoblasts (TB), stromal, Hofbauer, endothelial and nucleated red blood cells (nRBC)[42]. They observed that the most distinct DNAm profiles were those of placental TB, which presented the most pronounced placenta-specific methylation marks and were central to many essential functions. For instance, during gestation undifferentiated TBs change into fully differentiated STBs, forming a continuous, multinucleated, and specialized layer of epithelial cells[43]. In term placentas STB is the most abundant cell type with many key features as it covers the entire surface of the villous trees (Figure 1) and is in direct contact with maternal blood[44]. STBs orchestrate the complex biomolecular interaction between mother and fetus, and act as an endocrine organ, producing numerous growth factors and hormones that support and regulate placental and fetal development and growth[45-48]. More recently, a study published a few months ago demonstrated the significant dynamic heterogeneity of the STB nuclei, identifying their differentiation routes and their potential roles during early and late pregnancy[49]. But few studies or none have specifically investigated the role of cell type-specific epigenome in response to pregnancy exposures and/or leading to diverse health outcomes. The important role of placenta and the unique DNAm profile observed at both bulk tissue and cell type levels, lead us to speculate that the genome-environment interaction and its impact on DNAm is highly likely unique in placenta and deserves further investigation.

In the last decade, two of the most widely used analyses to study the epigenome are epigenome-wide association studies (EWAS) and methylation quantitative trait loci (mQTL).

2.1. Epigenome-wide association studies

The aim of EWAS is to examine the genome-wide epigenetic variation, predominantly DNAm at CpG sites, to detect differences that are significantly associated with phenotypes of interest. The most common way to study DNAm is with bisulfite converted genomic DNA and methylation arrays. Methylation arrays are the most affordable technology to assess genome-wide DNAm. Specifically, they are a collection of different oligonucleotides (probes) fixed on a solid substrate that can hybridize to complementary DNA strands. The DNAm levels are measured at each CpG site present on the microarray and compared between groups of interest to detect differentially methylated positions (DMP) and regions (DMR). A DMP is a single CpG dinucleotide that is differentially methylated between groups, as determined by statistical significance thresholds. The definition of a DMR differs between studies based on the algorithm used but can broadly be defined as a region containing multiple DMP[50]. Nowadays, two of the most commonly used microarrays are HumanMethylation450 (450K) and HumanMethylation850 (EPIC) that measure DNAm at over 450,000 and 850,000 CpG sites, respectively[51,52]. Illumina microarrays predominantly measure DNAm in gene promoter regions.

EWAS can be conducted analyzing unrelated case-control or longitudinal studies, as well as quantitative and qualitative phenotypes. A case-control EWAS is the most employed study design as it is more feasible and affordable in terms of organization and cost, compared to a longitudinal study. The case-control design is a standard design in epidemiology and involves grouping unrelated participants according to a phenotype of interest such as the presence/absence of disease. It compares CpG DNAm between that particular group and another without the phenotype[50] (Figure 3). It is important that control group matches for potential confounding factors such as age or sex with the case group.

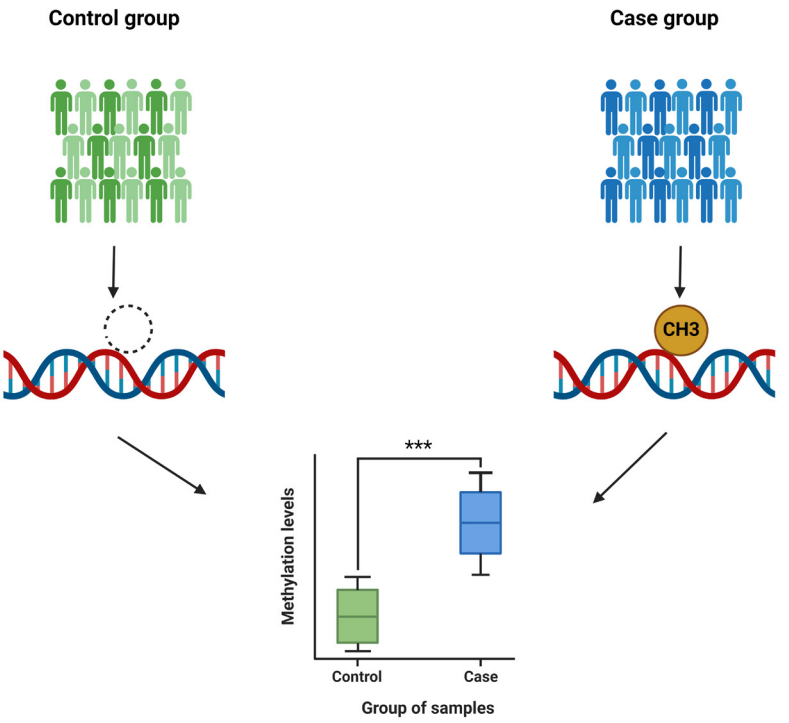


Figure 3. Schematic representation of a case-control EWAS study.

A case-control EWAS study compares CpG DNAm between two groups of unrelated samples. The group with the disease or condition under study is called the case group, otherwise the one without the condition is called the control group. A boxplot comparing the DNAm levels distribution between the two groups is used to illustrate the example depicted in the figure.

2.1.1. Pregnancy And Childhood Epigenetics consortium studies

In the last several years the Pregnancy And Childhood Epigenetics (PACE) consortium[53] from the National Institute of Health has led several EWAS that demonstrate that cord blood DNAm is sensitive to environmental factors surrounding gestation. The first PACE project was in 2016 and consisted in a meta-analysis of maternal smoking in relation to cord blood and child peripheral blood DNAm in newborns and children, measured using the 450K microarray. They identified more than 6,000 differentially methylated CpG sites in newborns in relation to maternal smoking during pregnancy. Out of this total, nearly 3,000 CpG sites, that correspond to 2,017 genes, had not been previously associated with smoking in either newborns or adults. Some of these genes have been identified in genetic studies of orofacial clefts, asthma, and lung and

colorectal cancer. Notably, many of these signals persist into childhood[54,55]. Nowadays, 39 cohorts are part of the PACE consortium and since 2016 many other exposures have been studied in cord blood and child peripheral blood DNAm, such as traffic related air pollution[56], maternal pre-pregnancy body mass index (ppBMI)[57], maternal alcohol intake[58], maternal C-reactive protein levels[59], maternal hemoglobin[60], but also, childhood health outcomes such as childhood asthma[61], BW[62], body mass index (BMI)[63] and attention-deficit and hyperactivity disorder (ADHD)[64].

Several PACE cohorts have also examined placenta DNAm demonstrating that this tissue is also sensitive to environmental factors surrounding gestation. In the first publication from 2021, a meta-analysis on 1,700 placental samples found a placenta-specific DNAm signature of maternal smoking during pregnancy (MSDP), with differentially methylated CpG sites located in active regions of the placental epigenome[65]. In total they identified 443 CpG sites that are associated with MSDP, of which 142 associated with birth outcomes (i.e. BW and birth length), 40 associated with gene expression, and 13 CpG sites were associated with all three. For example, decreased placental DNAm at cg27402634 correlated with increased expression of *LEKR1*, and associated with smaller BW and birth length. These placental MSDP-associated CpG sites were enriched for environmental response genes, growth-factor signaling and inflammation, all of which play important roles in placental function. Additionally, this DNAm signature was more marked for sustained smoking across pregnancy than for occasional smoking. And interestingly, by comparing the findings of this study with the ones obtained in the abovementioned 2016 study in cord blood, they found a minimal overlap between sites differentially methylated in the two tissues[54,65]. This observation highlights the tissue-specificity of the DNAm signatures, and thus, the importance of conducting the mentioned studies in placenta too.

Besides, maternal ppBMI is of particular interest since higher maternal ppBMI is associated with aberrant fetal growth[66], macrosomia[67] and increased neonatal morbidity, as well as with pregnancy complications, including preeclampsia, gestational diabetes, gestational hypertension, pre-term delivery and caesarean section[68]. It has been shown that maternal adipokine and insulin signaling in the placenta could contribute to regulate both the vascular development of this organ and the nutrient transport to the fetus, and therefore impact fetal development[68].

Additionally, it has been observed that maternal ppBMI is associated with other offspring health outcomes in later life, including increased risk for obesity in children[69]. Observational studies have suggested links between maternal obesity and long-term risk of coronary heart disease, stroke, type 2 diabetes, and asthma in offspring[70]. Very high maternal ppBMI has also been associated with poorer cognitive performance in children and greater risk of neurodevelopmental disorder[71], while there is also evidence in favor of potential implications in respiratory, immune and infectious disease-related outcomes[68,72-74].

The 2017 cord blood study from the PACE consortium showed that maternal ppBMI is widely associated with the DNAm at this particular tissue[57]. However, the authors observed that many significant epigenetic effects were modest (< 0.2% methylation per BMI unit) and they did not detect enrichments for any particular biological pathway, leaving open questions regarding the potential intra-uterine mechanisms that could be affecting the epigenetic profile of the newborn[57]. In this context, while the epigenetic alterations in cord and peripheral blood have been thoroughly investigated[57,75], the potential impact of maternal ppBMI in placental DNAm remains poorly explored. As far as we know, the most recent studies have performed methylation profiling with methylation arrays or reduced representation bisulfite sequencing in up to 300 term placentas of obese and non-obese mothers[76,77]. Although interesting, these studies have yielded a limited number of significant results, probably because of their relatively small sample size.

2.2. Methylation quantitative trait loci

mQTLs are single-nucleotide polymorphisms (SNP) whose genetic variation is associated with the DNAm levels of a CpG dinucleotide. We can categorize mQTLs as *cis*- and *trans*-mQTLs. On the one hand, *cis*-mQTLs are SNPs within a short distance on either side of the CpG position. On the other hand, *trans*-mQTLs are SNPs located at longer distance, even in a different chromosome, of the CpG position. Along this dissertation, when *trans*-QTL is not specified, the QTL is considered *cis*-QTL.

Other molecular quantitative trait loci (molQTL) have also been used to study the genetic determination of gene expression (eQTL), protein expression, splicing, RNA

editing, chromatin accessibility (caQTL) or histone modifications. MolQTLs, including mQTLs, aim to understand the functional effects of genetic variants, many of which are located in non-coding regions and associated with numerous traits and diseases[78]. In this sense, the most common genetic variants that impact human traits are believed to exert their effects through the regulation of gene expression[79,80]. Thus, in contrast to mQTLs, this molecular trait has been comprehensively characterized across many human tissues by the GTEx (Genotype-Tissue Expression) project, and eQTLs appear to underlie a substantial fraction of variant-trait associations[81,82]. However, our understanding of the regulatory mechanism by which variants influence human traits is far from complete, and elucidating how variants impact epigenetic factures such as DNAm is critical, since these features can influence, and respond to, gene expression. Indeed, the integration of mQTLs with genome-wide association study (GWAS) data has uncovered a putative role for DNAm in genetic regulatory mechanisms[83-88], such as the 92 putatively causal CpG sites for cardiovascular disease traits found in blood by Huan and colleagues in 2019[89]. More widely, Oliva *et al.* 2021 characterized the genome-wide DNAm profile of diverse, healthy human tissue types, and they observed that the aggregated contribution of mQTLs to human traits, in terms of number of identifiable associations, is larger than the one shown by eQTLs. Consequently, they demonstrated that mQTLs can reveal a substantial number of molecular links to traits otherwise missed by eQTL approaches, pinpointing putative candidate genes[90]. In this same study, they proved the importance of the mQTL catalogues derived from a variety of healthy, solid tissues to contribute to the characterization of the etiology of complex traits, given the existing differences in DNAm profiles across diverse tissues.

In this context and as we have already mentioned in point *2. Epigenetics, the bridge between genetics and environmental factors*, DNAm profiles are not only tissue-specific, but also cell type-specific. Most of the molQTL studies have been performed by using heterogeneous bulk tissue samples comprising diverse cell types. This limits the power, interpretation, and downstream applications of quantitative trait locus (QTL) studies. Genetic effects that are active only in rare cell types within a sample tissue may be undetected. Additionally, a mechanistic interpretation of QTL sharing across tissues and other contexts is complicated without understanding differences in cell type composition, and inference of downstream molecular effects of regulatory

variants without the specific cell type context is challenging[91]. Finally, efforts to map QTLs in individual cell types have been largely restricted to eQTLs and blood, using purified cell types[92-96] or single-cell sequencing[97]. However, single-cell and single-nucleus sequencing are not yet scalable to sample sizes and coverages sufficient to achieve a power comparable to that of bulk QTL studies[98-100]. Nevertheless, Kim-Hellmuth *et al*. 2020[91] and previous publications[101-103] proposed that cell type-specific QTLs can be computationally inferred from bulk tissue measurements by using estimated proportions or enrichments of relevant cell types to test for interactions with genotype. Kim-Hellmuth and colleagues found that cell type-interacting eQTLs (ieQTL) were strongly enriched for tissue and cellular specificity and provided a finer resolution to tissue specificity than that of bulk QTLs. Remarkably, given the enrichment of GWAS signals in cell type-ieQTLs for cell types potentially relevant to the traits, and the large fraction of colocalization with GWAS traits that are only found with cell type-ieQTLs, these new interacting QTLs seem to be highly valuable for gaining a mechanistic understanding of complex trait associations[91].

2.2.1. Placental mQTL studies

In the last few years there have been several studies that have mapped placental mQTLs. The first one is the one by Do *et al*. In this study 866 placental mQTLs were identified in 37 placenta samples. Some of the mQTLs detected had one or more SNPs in a 150 kb window associated with several diseases or traits, including body weight or schizophrenia (SCZ). However, this study was focused on multiple tissues and cell types, including brain, T lymphocytes, and purified neurons and glia. Thus, this interesting work lacked an extensive analysis of the placental mQTLs[104]. Later, in 2018, Delahaye and colleagues mapped mQTLs and eQTLs in 303 and 80 placentas, respectively, to clarify the interactions of genetic, epigenetic, and transcriptional regulatory mechanisms in human placentas[105]. In addition, they also mapped placental expression quantitative trait methylation (eQTM) in another set of 74 placental samples. eQTMs are CpG positions whose DNAm levels are associated with the expression of a nearby gene. This study detected 4,342, 1,916 and 2,566 mQTLs, eQTLs and eQTMs, respectively. Interestingly, they observed that the mQTLs were associated to CpG sites with intermediate DNAm

values, and their closest genes, mainly located in the major histocompatibility complex, were enriched for inflammation related pathways. This enrichment suggested that the variability captured by their associations may, in fact, reflect different degrees of inflammatory responses across our collected samples. Furthermore, to identify potential mechanisms underlying the correlation between DNAm and expression, they assessed the role of TF, looking at enrichment or specific binding sites overlapping CpG site from the eQTM associations. Associations with negative correlation were found to be globally enriched for TF binding sites suggesting a general mechanism where increase DNAm at a given binding site will alter binding abilities and affect transcription. As a matter of interest, associations that showed a positive correlation between expression and DNAm presented an enrichment for *ZNF217* TF, which has been previously recognized as a human oncogene and is known to be part of a complex that contains several histone-modifying enzymes strongly associated with gene expression[106]. Finally, they concluded that placenta is a highly heterogeneous tissue with particular profiles of gene expression and DNAm, which may result from changes in the proportions of different cell types that can later confound methylation-disease associations[105]. Nevertheless, both Do *et al*. 2016 and Delahaye *et al*. 2018 measured the methylation data on 450K array with a limiting number of DNAm sites that nowadays can be studied. At the same time, they analyzed the overlap between the genetic variants listed in GWAS studies and those associated with their mapped QTLs, instead of performing causality analyses. This limited the power to infer causality between the methylation or expression levels at specific genomic loci and different diseases and complex traits.

The subsequent study related to placental mQTLs is the one by Tekola-Ayele and colleagues, whose goal was to provide functional mechanistic insight into the causal pathway from genetic variants to BW by integrating placental methylation and gene expression with previously identified GWAS loci for this trait[107]. Considering BW GWAS lead SNPs discovered in Warrington *et al.*[108], Tekola-Ayele and colleagues mapped placental mQTLs and eQTLs in 291 and 71 placenta samples, respectively. Following, they conducted several causality analyses such as mediation, Mendelian Randomization (MR), and multiple-trait colocalization, with the loci to be both placental mQTL and eQTL. They found 813 and 32 mQTLs and eQTLs, respectively. Additionally, with the causality analyses they observe that methylation causally

influences genes *WNT3A*, *CTDNEP1*, and *RANBP2* expression in placenta, and concluded that their findings revealed candidate functional pathways that underpin the genetic regulation of BW via placental epigenetic and transcriptomic mechanisms[107]. However, their mQTL catalogue was limited to 450K methylation array and BW related loci, thus, limiting the application to other traits.

More recently, Casazza *et al*. analyzed 411 placenta samples to study whether placental mQTLs explain part of the genetic risk for childhood-onset traits and if so, whether it acts in a sex-dependent manner[109]. Therefore, they mapped four different placental mQTL sets: cross-sex mQTLs or the standard placental mQTLs, sex-interacting mQTLs or those in which the effect of SNP on DNAm differed by sex, and mQTLs identified only on males or females. They found 49,252 cross-sex mQTLs and 2,489 sex-dependent mQTLs. In the case of the sex-dependent mQTLs, they mapped 351 male-specific mQTLs, and 255 female-specific mQTLs. Of the 351 male-specific mQTLs, 185 (35%) also had a cross-sex mQTL. Of the 255 female-specific mQTLs, 153 (60%) also had a cross-sex mQTL. 75 CpG sites had both male- and female-specific mQTL effects, of which 74 (99%) also had a cross-sex mQTL. Briefly, they noticed that placental mQTLs were enriched in regions controlling gene expression and primarily occurred at CpG sites with intermediate levels of DNAm. Interestingly, they spotted that sex-dependent and cross-sex mQTLs showed similar patterns of tissue specificity compared to two other prenatal tissues, umbilical cord and fetal brain. To study how placental mQTLs contribute to the genetic risk of childhood traits and conditions, they performed a stratified-linkage disequilibrium (LD) score regression analysis to estimate the proportion of SNP-heritability of 19 complex traits that was explained by their placental mQTL sets. They noticed that all mQTL sets were enriched in GWAS results from growth- and immune-related traits, but male- and female-specific mQTLs were more enriched than cross-sex mQTLs[109]. Even though the evident potential of these mQTLs sets, the complete mQTL databases remain unpublished and thus inaccessible to the scientific community for further studies.

Considering all these studies, we conclude that there is no publicly accessible placental mQTL database measuring DNAm on more other than those on the 450K array, including a considerable sample size, and considering cell type heterogeneity and its specific epigenetic profiles.

3. Colocalization and Mendelian Randomization analyses

Nowadays, GWAS have identified thousands of genetic variants associated with human complex traits. However, the genes or functional DNA elements through which these variants exert their effects on the traits remain unknown, and as it has already been mentioned, the overlap between genetic variants from GWAS and QTL studies do not imply causality. Inferring causal relationships between phenotypes is a major challenge and has important implications for understanding the etiology of health and disease. Additional approaches such as colocalization and MR analyses can be useful to statistically infer the causal effects between two traits, namely a molecular trait (i.e. DNAm or gene expression) and a disease.

Colocalization analyses test whether two independent association signals at the same locus are consistent with a shared causal variant. If the answer is positive, we refer to this situation as colocalized traits, and the probability that both traits share a causal mechanism greatly increases. A common example involves an eQTL study and a disease association result, which point to the causal gene and the tissue in which it exerts its effect on the disease[110-112]. Another example, although less frequent, applies to mQTL studies where a DNAm site is pinpointed instead of a gene[84]. One of the most used colocalization tools is the Coloc R package. For instance, Coloc is a Bayesian statistical procedure that assesses whether two association signals colocalize, thus, it tests whether a shared causal variant might be causing both traits, for example, DNAm at a given site and a disease[113]. Briefly, Coloc evaluates five different hypotheses (Figure 4). Hypothesis H_0, H_1 and H_2 correspond to situations without causal SNPs in both DNAm and disease, H_3 to a situation where DNAm and disease have different causal SNPs, and H_4 where DNAm and disease have consistent evidence of a shared causal SNP, i.e., colocalization (see the *Methods* section for more details). Since its publication in 2014 several studies have been performed using this method. For example, *TACSTD2* has been described as a novel therapeutic target for

cisplatin-induced acute kidney injury, a common complication in cancer patients[114]. And similarly, altered DNAm within *DNMT3A, AHRR* and *LTA/TNF* loci have been identified as mediators of the effect of smoking on inflammatory bowel disease[115].

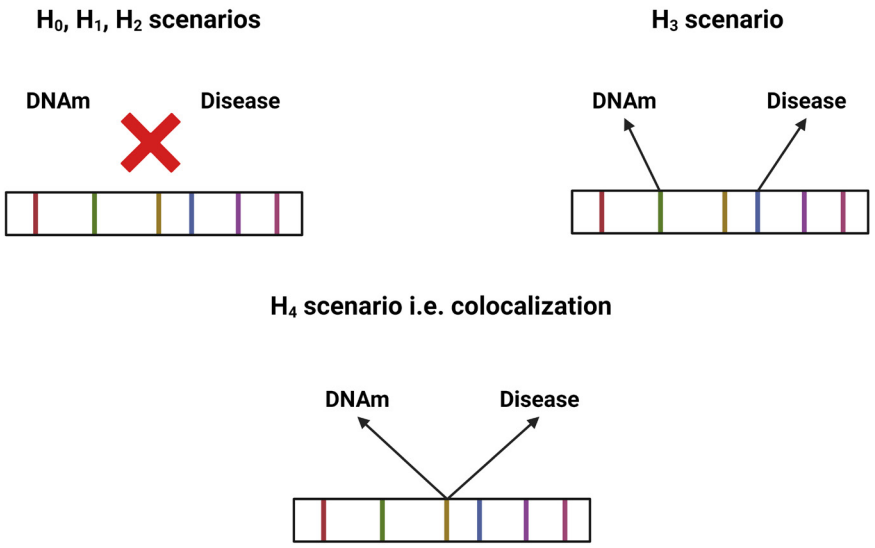


Figure 4. Schematic representation of the five hypotheses considered in colocalization analyses.

Each color bar represents an SNP. Hypothesis H_0, H_1 and H_2 correspond to situations without causal SNPs in both DNAm and disease, H_3 to a situation where DNAm and disease have different causal SNPs, and H_4 where DNAm and disease have consistent evidence of a shared causal SNP, i.e., colocalization.

Besides, MR uses genetic variation to address causal questions about how modifiable exposures influence different outcomes. The principles of MR are based on Mendel's laws of inheritance and instrumental variable (IV) estimation methods, which enable the inference of causal effects in the presence of unobserved confounding[116]. Let us suppose that we have a SNP that is known to influence a given phenotype (the exposure); considering Mendel's laws and the fixed nature of germline genotypes, the alleles that an individual receives at this SNP are expected to be random with respect to potential confounders and causally upstream of the exposure. In this 'natural experiment' the SNP is considered to be an IV and observing an individual's genotype at this SNP is akin to randomly assigning an individual treatment or control group in a randomized control trial. Thus, MR uses

genetic variation to mimic the design of randomized control trial. To infer causal influence of the exposure, one calculates the ratio between the SNP effect on the outcome over the SNP effect on the exposure[117]. Most of the MR studies use the two-sample MR design, in which one cohort of subjects has measurements for the exposure, while a second cohort of subjects has measurement for the outcome, with both cohorts sharing the same set of genetic variants[118]. This enables us to use summary data from GWAS and QTL studies to harness the statistical power of pre-existing analyses. Due to the flexibility of this two-sample strategy, MR can be applied to thousands of potential exposure-outcome associations, where the exposure can be very broadly defined, from gene expression to DNAm, to more complex traits, such as BMI or intestinal microbiota.

While MR avoids certain problems of conventional observational studies[117], it introduces its own set of problems. MR is predicated on exploiting vertical pleiotropy (i.e. causal scenario), where a SNP influences two traits because one trait causes the other[119]. It is crucial to be aware of the assumptions and limitations that arise due to this model[120]. The main assumptions are (Figure 5A): the instrument associates with the exposure (IV assumption 1); the instrument does not influence the outcome trough some pathway other than the exposure (IV assumption 2); and the instrument does not associate with confounders (IV assumption 3)[117]. While assumption IV1 is easily satisfied in MR by restricting the instruments to genetic variants that are discovered using genome-wide levels of statistical significance, the other two assumptions are impossible to prove, and when violated, can lead to bias in MR and introduce horizontal pleiotropy. This type of pleiotropy takes place when the SNP influences the outcome through some pathway other than the exposure.

When molecular traits are considered as the exposure, the most used MR tool is the Summary-based MR (SMR) software. SMR uses summary data from QTL and GWAS studies to test for pleiotropic association between the expression of a gene and a complex trait of interest using summary-level data from eQTL studies and GWAS[121]. The SMR methodology can be interpreted as an analysis to test if the effect size of a SNP on the phenotype is mediated by gene expression. This tool can therefore be used to prioritize genes underlying GWAS hits for follow-up functional studies.

Although SMR was originally developed to test gene expression as the exposure, the method is applicable to all kinds of molecular QTL data, including mQTLs. Thus, it allows you to prioritize CpG sites whose DNAm is mediating the SNP effect on a phenotype. In SMR three possible explanations for an observed association between the outcome and the exposure through genotypes are considered (Figure 5B): the abovementioned vertical and horizontal pleiotropy, and linkage. For instance, linkage is when the association observed is due to the top associated QTL being in LD with two distinct causal variants, one affecting the exposure and the other affecting the outcome. Although SMR is not able to distinguish vertical from horizontal pleiotropy, hence having to complement the results alongside with additional analysis, it can differentiate linkage from pleiotropy. To do so, SMR test is complemented by the HEterogeneity In Dependent Instruments (HEIDI) test in which multiple SNPs in QTL regions are used to test for the heterogeneity of the effect calculated with the different IVs. If heterogeneity is detected, we cannot rule out the possibility that two different variants in LD are affecting independently the exposure and the outcome[121] (see the *Methods* section for more details).

In the last few years, a wide range of studies have used SMR to point out causal genes or genomic loci to understand the etiology of a disease by integrating several omics. Studies such as Liu *et al.* 2022, Rahmioglu *et al.* 2023 and Xu *et al.* 2023 integrated mQTLs, eQTLs, caQTLs and fecal microbial QTLs to study the origin of endometriosis, kidney disease and Crohn's disease with the final aim of providing evidence for future targeted functional research aimed at developing suitable therapeutic interventions and disease prevention[122-124]. In the case of the endometriosis study, several potentially causal genes were reported, including *GREB1*, *VEZT* and *FGD6*, confirming previous evidence. Their related risk variants may function through changes in DNAm and expression patterns in blood and endometrium[122].

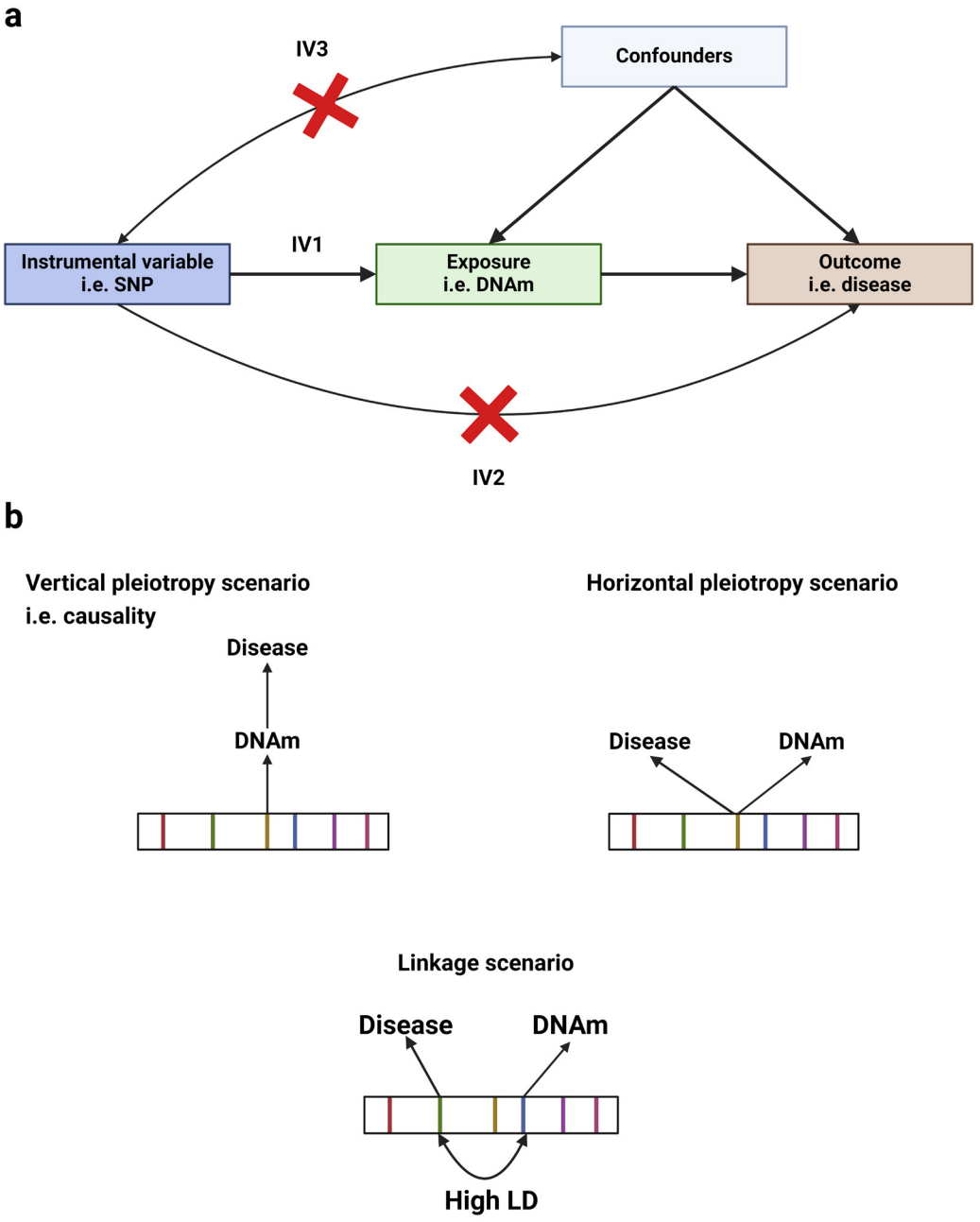


Figure 5. Schematic representation of Mendelian Randomization.

In **a**, the three MR assumptions are represented, and it has been adapted from Hemani *et al.* 2018[117], and in **b**, the three SMR scenarios are depicted, and it has been adapted from Zhu *et al.* 2016[121]. Each color bar represents an SNP.

4. The association between neuropsychiatric traits and the prenatal environment

Neuropsychiatric conditions that may not manifest until childhood or young adulthood are increasingly being recognized to have their origins in the fetal period. As such, maternal health and intrauterine environmental exposures can influence fetal brain development, often through converging physiological pathways. Some of these adverse exposures include maternal infection and inflammation, hypoxia, polysubstance use, maternal stress, anxiety, and depression, as well as placental insufficiency and dysfunction. Of note, throughout this dissertation, when the term depression is used, it refers to depression symptoms, not to major depressive disorder (MDD) which requires a particular clinical diagnosis and in fact is the severe form of depression.

In the case of maternal infection and inflammation, both placenta and immature fetal blood-brain barrier have increased permeability to toxic and inflammatory mediators, many of which can compromise proper growth and development of the fetus[125]. The pathways by which this exposure might be acting are unclear, but some hypotheses point to 1) a direct effect of antibodies or toxins on neurodevelopment, 2) activation of immune response and mediators, and 3) the interaction between environmental factors and pre-existing genetic vulnerabilities. For instance, maternal immune activation (MIA), which is often defined as the maternal exposure to, or infection with, various immunogens during pregnancy, has been associated with increased risk for ASD and SCZ in offspring in several epidemiological studies[126-128] and animal models[129,130].

In animal models, chronic hypoxia in the setting of placental insufficiency or other obstetric complications has been proposed to play a role in impaired fetal neurodevelopment, as decreased oxygen delivery in the fetal brain has been associated with later disruptions of the serotonin and noradrenaline systems[131-133].

These neurotransmitters are crucial in the developing brain, fundamentally for the normal function of memory, learning, behavior and movement. Alterations in these functions have been described in several neuropsychiatric disorders, such as SCZ, ASD and ADHD[134-136].

Maternal use of alcohol, cigarettes, marijuana, illicit drugs and opioids during pregnancy can adversely impact fetal mental health through direct effects causing fetal abnormalities or influencing brain biochemistry through alterations in neurotransmitters or neurotrophins[137,138]. It is known that nicotine, cocaine, methamphetamine and high levels of alcohol consumption have several adverse effects, such as the alteration and activation of neurotransmitter pathways, as well as the modification of brain structure and volume, with elder children exhibiting attention and inhibitory control deficits[139-143]. For example, a study conducted in Hawaii showed that many birth defects are more prevalent among children exposed to cannabis during gestation[144]. Another study observed that young school-aged children exposed to methamphetamine during pregnancy exhibit anxiety, emotional instability, aggression (AGR), and personality disorders such as ADHD[145].

Maternal mental health conditions can further influence fetal brain and have been associated with several neuropsychiatric conditions. Maternal stress, anxiety and depression have been associated with deviations in neurodevelopment and with neuropsychiatric conditions, including emotional problems, ADHD, impaired cognitive development, or ASD[146]. These presumably occur through alterations in inflammatory, neurotransmitter-related and neuroendocrine pathways[147]. Maternal stress hormones are known to cross the placenta, simulating the fetal stress axis (hypothalamic-pituitary-adrenal, or HPA axis) and releasing fetal cortisol. Early fetal exposure to these stress hormones is associated with preterm delivery, long-term metabolic effects on the child, including obesity and hypertension, and altered behavioral and cognitive functions[148-150]. Alterations of the HPA axis are also associated with mood disorders in adulthood, including MDD[151].

As we have already mentioned in the section *1.1. The role of placenta in DOHaD*, the placenta is a key organ for a healthy development, and its dysfunction, including

insufficient nutrient supply, periods of hypoxia, decreased protection from pro-inflammatory cytokines and altered placental metabolism, is one of the most significant risk factors for neuropsychiatric diseases[152]. These placental conditions could arise not only from environmental insults but also from a given genetic susceptibility background and therefore deserve further investigation. Some common neuropsychiatric disorders are ASD, ADHD, SCZ and bipolar disorder (BIP).

4.1. Autism spectrum disorder

ASD is a neurodevelopmental disorder that impairs normal brain development and socio-cognitive abilities[153]. Several previous studies have related prenatal factors and increased risk for ASD. For example, a meta-analysis of 15 studies demonstrated that offspring from mothers who had acquired infections during pregnancy had up to 1.3-fold increased risk for developing ASD[127]. Another meta-analysis performed in 2015 reported a two-fold increased risk for children exposed to preeclampsia[154]. But in the last years, it has been proposed that genetic and environmental factors during life *in utero* can be mediated by epigenetic modifications in placenta. In relation to these, several EWAS have identified thousands of differentially methylated CpG sites in the placenta of future ASD patients[155-159]. DMRs in the ASD-placentas are dispersed throughout the genome[157] and have been identified to affect biological pathways converging on synaptogenesis, microtubule dynamics, neurogenesis, and neuritogenesis, which finally influence neuron morphology, brain development and cognitive abilities[155,156]. Some of the most robustly differentially methylated genes reported in the different studies include *DLL1*, *NEUROG2* and *POU3F2*. For instance, *DLL1* has been reported to be hypermethylated in the placenta of ASD[156,160], and influences brain development by regulating neurogenesis and neuron differentiation[161,162]. *NEUROG2* has been found to be hypomethylated in ASD[156]. It induces differentiation and survival of midbrain dopaminergic neurons, and its knockout in cell cultures results in lack of this type of neurons, demonstrating a potentially relevant function in ASD[163,164]. Finally, *POU3F2*, which has been described to be hypomethylated in ASD placentas[156], encodes a TF involved in neuronal differentiation and is considered a master regulator of downstream ASD candidate genes[156,165].

4.2. Internalizing and externalizing behavior

Childhood behavioral and emotional problems are associated with increased risk for a wide-spectrum of adverse outcomes during adolescence and adulthood, such as adolescence aggression and adult violence[166-171]. Indeed, the Global Burden of Disease study from 2010 estimated that behavioral disorders in childhood accounted for nearly 6 million disability-adjusted life years[172], rendering childhood behavioral problems an increasingly critical public health concern and major social issue. Internalizing (INT) and externalizing behaviors are a well-established and widely used behavioral classification within the field of child and adolescent psychology[173]. Internalizing problems refer to inwardly focused negative behaviors such as anxiety, depression and somatic symptoms[174], as well as different disorders, including obsessive-compulsive disorder (OCD). Externalizing problems refer to outwardly focused negative behaviors such as AGR, disruptive conduct and several disorders, including ADHD[175,176]. Several prenatal risk factors, including health-related and psychosocial factors, have been studied in this context and have been reported to have an effect not only in childhood, but also in later stages such as adolescence[177-179].

An Australian cohort found that increased maternal ppBMI was significantly associated with INT problems that emerged at age 8 years and increased until age 17 years[177]. Regarding MSDP, a Dutch twin-children study observed a direct causal effect of MSDP on externalizing behavior at age 3 years, with mothers' smoking cessation prior to conception associated with less overall externalizing problems, including AGR and oppositional behavior[180]. In terms of INT, the Norwegian Mother and Child Cohort Study, also known as MoBA, reported that maternal smoking during early pregnancy shared a dose-response relationship with increased anxiety and depression behaviors in children at 18 months and 3 years of age[181].

In relation to alcohol consumption, the Avon Longitudinal Study of Parents and Children, also known as ALSPAC, found that episodic binge drinking patterns are associated with externalizing behaviors in 11 year-old children[178]. Regarding prenatal MIA, Giollabhui *et al.* found that maternal inflammation during pregnancy predicts

more severe offspring internalizing symptoms in childhood, with especially marked effects noted in female offspring[182]. During adolescence, Murphy *et al.* found that maternal infection, specifically during the second trimester, was significantly associated with offspring depression at ages 15-17 years[179].

Finally, there have been studies on psychosocial factors such as intimate partner violence (IPV), which has been declared as a global public health concern by the World Health Organization[183]. Current studies reported significant associations between prenatal IPV and both externalizing and INT behavior, especially anxiety and depression, across infancy at 12 and 24 months of age, respectively[184,185].

In terms of placental DNAm and internalizing and externalizing behaviors, Nakamura *et al.* found three DMPs and 33 DMRs associated with at least one of the Strengths and Difficulties Questionnaire subscales assessed at 3 years of age in 441 mother-child dyads from the study on the pre- and early postnatal determinants of child health and development (EDEN - *Étude de cohorte généraliste menée en France sur les Déterminants pré- et postnatals précoces du développement psychomoteur et de la santé de l'ENfant*)[186]. This study provided the first evidence of association between placental DNAm and child behavioral and emotional difficulties.

Up to date, most of the studies have analyzed depression as a symptom of INT behavior, but none of them has analyzed the effect of prenatal stages on the development of MDD in adulthood.

4.2.1. Attention and deficit hyperactivity disorder

ADHD is an externalizing disorder and a psychiatric condition that presents patterns of developmentally inappropriate levels of inattentiveness, hyperactivity, or impulsivity[187]. Unlike ASD, conflicting data have been revealed to identify prenatal risk factors for ADHD development. In this pathology, the association between maternal infection and ADHD in offspring is unclear. A recent study of 114,000 children demonstrated an elevated risk for ADHD in offspring exposed to maternal fever in the first trimester, with an increasing risk in the setting of multiple episodes of fever[188]. But other studies have found that this association with ADHD was not coming from

the fever i.e. maternal infection, but from the maternal use of acetaminophen, a medication to lower fever[189,190].

In terms of maternal lifestyle factors, systematic reviews suggested that nicotine and tobacco exposures increase the risk of ADHD 1.6-fold, but there are conflictive findings regarding alcohol exposure, environmental toxin exposures and maternal psychological stress[191]. Indeed, one of the strongest risk factors for the development of ADHD is preterm and low BW, with a three-fold increase in the risk of developing ADHD in very extreme preterm or low BW[192]. The underlying mechanisms are complex and may relate to placental insufficiency resulting in low BW and preterm delivery[192]. Although few studies have described epigenetic mechanisms associating prenatal conditions and ADHD development in cord blood[193,194], none has investigated the role of the placenta in the development of this specific disorder.

4.3. Schizophrenia and bipolar disorder

SCZ is a chronic, psychiatric disorder characterized by a wide variety of symptoms, including delusion, hallucinations, disorganized speech or behavior, and impaired cognitive ability. Additionally, the early onset of the disease, along with its chronic course, makes it a disabling disorder for many patients and their families[195]. Nowadays, numerous studies have investigated the contribution of prenatal environment to SCZ development. One of the first studies was performed in 1957, when Mednick *et al*. investigated the effects of the influenza pandemic on adults who were born during the epidemic versus those who were not, concluding that those that were exposed *in utero* to influenza were also more likely to develop the disorder in adulthood[196]. Afterwards, in 1987, Daniel Weinberg proposed the neurodevelopmental hypothesis of SCZ which has been reinforced in the last decades, with increasing evidence of neurodevelopmental abnormalities contributing to the pathophysiology of the disease[197-199]. The central argument of this hypothesis states that abnormal fetal neurodevelopment creates a vulnerability to develop SCZ later in life. In fact, different pieces of evidence state that prenatal insults such as MIA and other pregnancy- and labor-related triggers such as hemorrhage, preeclampsia, birth asphyxia, and

uterine rupture are associated with SCZ in offspring[200,201]. Specially, MIA occurs when inflammatory markers rise above the normal range in a pregnancy as a result of maternal inflammation, and can be caused by psychosocial stress, infection, or other factors[202]. Human and animal studies suggest that increased inflammatory cytokines, and genetic expression of cytokine receptors and the histocompatibility complex region are significant prenatal contributors in the etiology of SCZ[203,204]. Several genes such as *BDNF*, *COMT*, *GAD1, RELN* and *SOX1* have been shown to be differentially methylated in blood and brain tissues from SCZ patients[205-209], but no studies have been conducted in the placental epigenome. In relation to placenta, one significant study from Ursini and colleagues published in 2018 reported that the genomic risk for SCZ is up to five times higher in the presence of pregnancy complications, and that genes present in those risk loci are highly expressed in placenta and, specifically, in placenta from complicated pregnancies[210]. In a more recent study, the same team analyzed whether fractionated genomic risk scores for SCZ and other developmental disorders and traits, based on placental gene-expression loci, were linked with early neurodevelopmental outcomes in individuals with a history of early life complications (ELC). They found that placenta-specific SCZ genomic risk scores were negatively associated with neonatal brain volume both in singletons and in the offspring of multiple pregnancies, and with cognitive development in singletons at 1 year of age and, less strongly, at 2 years of age, when cognitive scores become more sensitive to other factors[211]. Additionally, they observed that the relationship of the placental gene-expression loci identified by Ursini *et al.* with brain volume persists as a basis of placental biology in adults with SCZ, selectively in males. They concluded that higher placental genomic risk for SCZ, in the presence of ELC and particularly in males, alters early brain growth and function, defining a potentially reversible neurodevelopmental path of risk that may be unique to SCZ[211].

BIP is a chronic, neuropsychiatric disorder characterized by episodes of mania and hypomania alternating with depression[212]. It is also the leading cause of disability in young people as it can lead to cognitive and functional impairment and increase mortality, particularly due to higher incidence of suicide and cardiovascular disease[213]. While its etiology is widely unknown, there has been some interest in the role of perinatal and prenatal risk factors in the development of this condition[214], mostly

considering that BIP shares many risk loci and pathways with SCZ[215]. However, up to date, very few studies have investigated perinatal insults in the context of BIP development. In 2014, perinatal asphyxia was associated with smaller left amygdala volume observed in BIP patients, but not in healthy controls[216]. Patients with psychotic BIP showed an association between perinatal asphyxia and smaller left amygdala volume, whereas patients with non-psychotic BIP showed smaller right hippocampal volumes related to both perinatal asphyxia and sever obstetric complications.

In conclusion, more studies are demonstrating the association between pregnancy exposures, such as maternal ppBMI, and the later development of a wide range of traits and disorders in the context of DOHaD hypothesis. Similarly, epigenetic studies have demonstrated to function as a bridge between genetics and environmental factors, but most of them have been carried on cord blood tissue, leaving aside the potential role of placenta as one of the key players in a successful pregnancy. In the last few years, novel studies have shown that this organ has specific and unique epigenetic marks, different from those identified in cord blood or any other tissue. These patterns might be one of the mechanisms by which prenatal exposures and susceptibility genes exert their effects. In this sense, some neuropsychiatric disorders have been suggested to have prenatal origins, but a limited number of articles have focused on the placental epigenome as the potential mechanism by which genetics might be contributing to the development of these disorders.

Thus, on the one hand, our first hypothesis is that the placental epigenome is sensitive to maternal ppBMI and shows a unique DNAm pattern not observed in other tissues. And, on the other hand, we hypothesize that the placental methylome could be a key mechanism by which environment and genetic risk interact during pregnancy, and therefore confers susceptibility to develop several neuropsychiatric disorders later in life.

AIMS

Aims

The overall aim of this thesis was to elucidate the impact of maternal ppBMI, a relevant fetal environmental factor, and fetal genetics on later life health via placental DNAm.

The operative aims of the present study were:

1. To determine the possible association of maternal ppBMI with epigenome-wide placental DNAm, and to understand the mechanisms by which maternal obesity could be associated with future health outcomes in offspring.
2. To construct and to make publicly available the largest placental *cis*-mQTL database to date, as a novel tool to understand the effect of placental DNAm in later life health.
    a. To identify and to characterize the placental DNAm sites regulated by genetics, and more specifically by SNPs.
    b. To identify and to characterize the placental DNAm sites regulated by the interactions between SNPs and cell type proportions, gestational age (GA), and fetal sex.
3. To ascertain whether placental DNAm is the mechanism by which genetic risk factors contribute to the development of several neuropsychiatric diseases, including ADHD, AGR, ASD, BIP, INT, MDD, OCD, panic disorder (PD), suicidal attempt (SA), and SCZ.

METHODS

1. Placenta biopsies and DNA extraction

The Environment and Childhood study (INMA - *INfancia y Medio Ambiente*)[217] is a population-based mother-child cohort study in Spain that aims to study the role of environmental pollutants in air, water and diet during pregnancy and early childhood in relation to child growth and development. More information about the INMA project is available through our webpage http://www.proyectoinma.org/. The present study is based on the four *de novo* cohorts from Asturias, Gipuzkoa, Sabadell and Valencia that were recruited between 2003 and 2008. The study was approved by the ethical committees of the centers involved in the study, and written informed consent was obtained from all the participants. In INMA 2,506 mother-fetus pairs were followed until birth and a selection of 489 placentas were collected and stored at -80 ºC in a central biobank until processing. Biopsies of approximately 5 cm^3 were obtained from the inner side of the placenta, approximately 1.0-1.5 cm below the fetal membranes, corresponding to the villous parenchyma, and at ≈5 cm from site of the umbilical cord insertion. 25 mg of placental tissue were used for DNA extraction, previously rinsed twice during 5 minutes in 0.8 mL of 0.5X PBS to remove traces of maternal blood. Genomic DNA from placenta was isolated using the DNAeasy® Blood and Tissue Kit (Qiagen, CA, USA). DNA quality was evaluated in a NanoDrop spectrophotomer (Thermo Scientific, Waltham, MA, USA) and additionally, 100 ng of DNA were run on 1.3% agarose gels to confirm that samples did not present visual signs of degradation. Isolated genomic DNA was stored at -20ºC until further processing.

2. Genotype data quality control

Genome-wide genotyping was performed using the Illumina GSA BeadChip at the Human Genotyping Facility (HuGeF), Dept Internal Medicine, Erasmus MC, Rotterdam, the Netherlands, and the Spanish National Genotyping Centre, CEGEN, Madrid, Spain. Genotype calling was done using the GeneTrain2.0 algorithm based on HapMap clusters implemented in the GenomeStudio software. Samples were genotyped in four batches.

The quality control of the genotype data from 397 INMA samples and 509,450 genetic variants was performed using the PLINK 1.9 software following standard recommendations[218-220]. Prior to imputation, PLINK files were processed with Will Rayner's preparation Perl script available from Mark McCarthy's Group as recommended in the documentation from the Michigan Imputation Server[221], using the Haplotype Reference Consortium (HRC) r1.1 2016 reference panel[222]. Variants with a call rate below 95%, minor allele frequency (MAF) below 1%, or a P value from the Hardy-Weinberg Equilibrium (HWE) exact test below $1x10^{-6}$ were removed. Samples with discordant sex, those with average heterozygosity values above or below 4 standard deviations (SD), or with more than 3% missing genotypes were filtered out. Identity-by-descent values were calculated with PLINK, and from those sample pairs that showed PI-HAT estimates above 0.18, the sample with higher proportion of missing genotypes was removed.

The final data set was imputed with the Michigan Imputation Server using the HRC reference panel, Version r.1.1 2016. Before imputation, data were converted into variant call format (VCF). Phasing of haplotypes was done with Eagle v2.4[223] and genotype imputation with Minimac4[224], both implemented in the code available at the Michigan Imputation Server. Finally, we removed variants with an imputation quality r^2 below 0.9, a MAF lower than 5%, a HWE P value below 0.05 and with more than two alleles, to avoid SNPs with few or no individuals bearing the minor allele homozygous genotype in our sample set. Only those samples with DNAm data were considered in this analysis. The final data set consisted of 368 samples and 4,171,035 SNPs. Figure 6 illustrates a diagram with the main steps followed by this quality control.

3. DNAm data quality control

DNAm from the fetal-facing side of 190 and 397 placentas was analyzed with the Illumina 450K and EPIC arrays for the maternal ppBMI EWAS and the placental mQTL database, respectively, in the Erasmus Medical Centre core facility, following the manufacturer's protocol. Briefly, 750 ng of DNA from 397 placental samples were bisulphite-converted using the EZ 96-DNAm kit from Zymo Research, following the manufacturer's standard protocol, and DNAm was measured using the Infinium protocol. Three technical duplicates were included. Samples were randomized considering region-of-origin and sex. As the number of samples in each condition was different, perfect randomization was not possible. However, all the plates had samples from all three geographical areas involved, and an equilibrated number of male and female samples.

3.1. DNAm data quality control of the 450K array

For the 450K array, the initial data set consisted of 190 samples and 509,450 DNAm probes or CpG sites. Low-quality samples (showing a shifted beta-value distribution) were filtered out and probes with detection P values > 0.01 were excluded. DNAm beta values were normalized with functional normalization, and beta-mixture quantile normalization (BMIQ)[225] was applied to correct for probe type bias. The batch effect was examined by depicting boxplots that divided the samples into separate groups according to suspicious variables, after ComBat had been applied[226]. Probes that hybridized with the X/Y chromosomes, cross-hybridizing probes, and probes with SNPs with an average MAF in all populations > 1% at the CpG site or in the probe region were filtered out[227]. Finally, DNAm extreme outliers (< 25th percentile - 3*IQR or > 75th percentile + 3*IQR across all the samples) were trimmed. The final data set consisted in 168 samples and 405,301 DNAm probes or CpG sites. Figure 6 illustrates a diagram with the main steps followed by this quality control.

3.2. DNAm data quality control of the EPIC array

For the EPIC array, the initial data set consisted of 397 samples and 865,859 DNAm probes or CpG sites. The quality control of the DNAm data was performed using the PACEAnalysis R package (v.0.1.7)[228]. Before the quality control with PACEAnalysis, one sample was discarded because of too many missing values in relevant variables. With the R package, we discarded those samples with a call rate below 95%, sex inconsistencies, intentioned or non-intentioned duplicates and those contaminated with DNA from another subject or the mother. Only those samples with genotype data were considered in this study. Probes with a call rate lower than 95%, a detection P value > 0.05, SNPs with an average MAF in the European population ≥ 5%, cross-hybridizing potential and that hybridize with X/Y chromosomes were excluded from the analysis[229].

The methylation beta values were normalized in different steps: firstly, dye-bias and Noob background correction, implemented in minfi R package, were applied[230,231], followed by normalization of the data with the functional normalization method[232]. Then, to correct for the bias of type-2 probe values, BMIQ normalization was applied[225]. After that, we explored the clustering of the data through Principal Component Analysis (PCA) and tested the association of the 12 first Principal Components (PC) with the main and the technical variables. Array batch effects were controlled with the ComBat method[226]. To correct for possible outliers, we Winsorized the extreme values to the 1st percentile (0.5% in each side), where percentiles were estimated with the empirical beta distribution. Finally, the rank-based inverse normal transformation (RNT) was applied to the beta values, and these estimations were the DNAm values considered for mQTL mapping. The final data set consisted of 368 samples and 747,486 DNAm probes or CpG sites. Figure 6 illustrates a diagram with the main steps followed by this quality control.

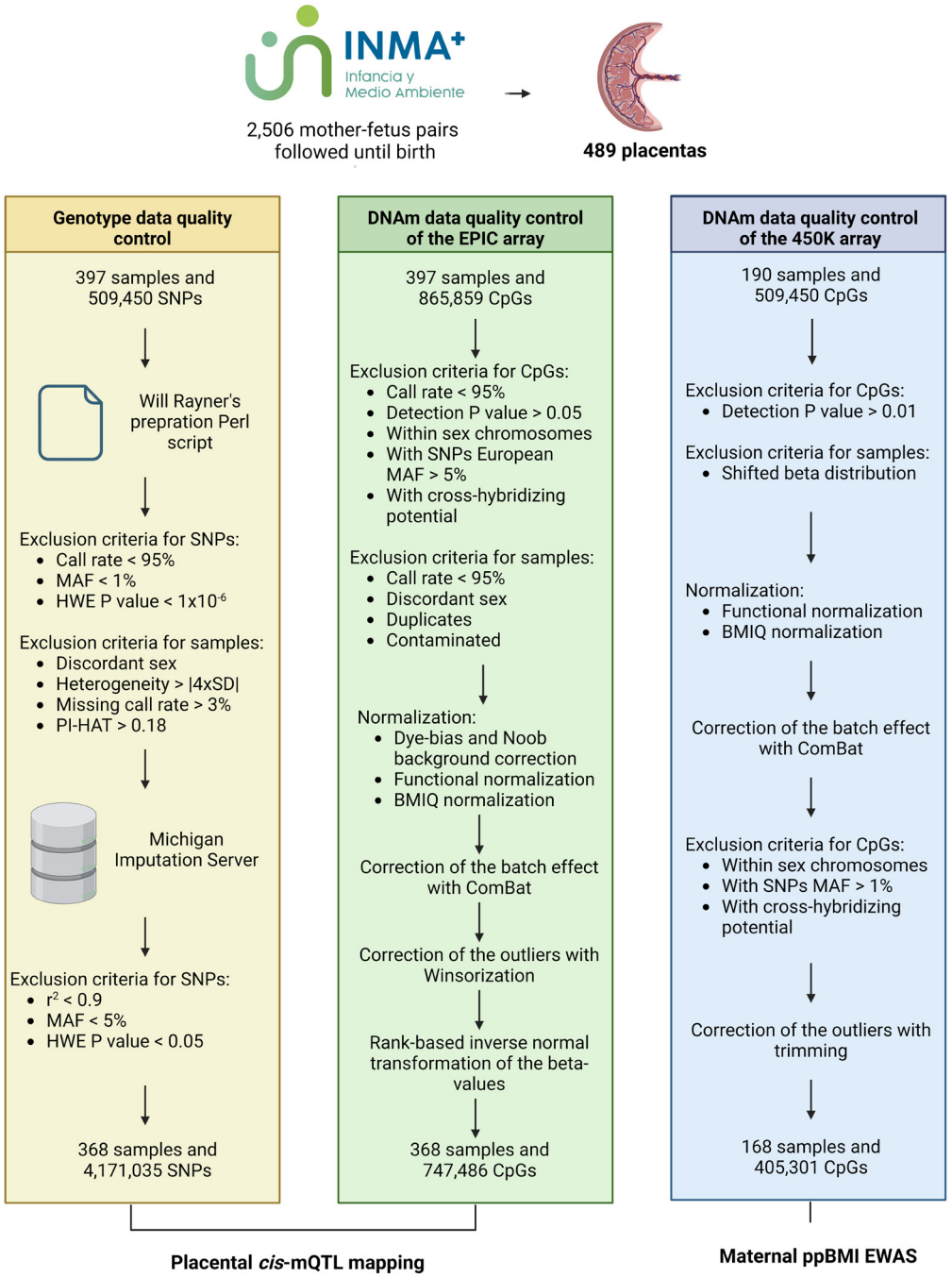


Figure 6. Diagram of the methodology followed by genotype and DNAm quality control from the INMA placental samples.

This detailed diagram illustrates the main filters and processes that genotype, DNAm from EPIC array and DNAm from 450K array data have followed in each of the studies performed. Additionally, the initial number of samples, SNPs, and CpG sites or probes is specified in each process.

4. Estimates of putative cellular heterogeneity

The cellular heterogeneity of the samples was estimated from DNAm data using a reference-free and a reference-based method for the maternal ppBMI EWAS and the placental mQTLs, respectively.

4.1. Reference-free method

Putative cellular heterogeneity was estimated from DNAm data using the reference-free cell-mixture deconvolution method RefFreeCellMix[233]. As a result, the number of components selected was 2. Models for differential DNAm were corrected for the number of surrogate variables minus one to reduce multi-collinearity.

4.2. Reference-based method

Putative cell type proportions of six populations (STB, TB, nRBC, Hofbauer cells, endothelial cells, and stromal cells) were estimated from DNAm data using the placenta reference panel from the 3rd trimester implemented in the Planet R package[42] using the constrained Houseman method. Mapping of placental mQTLs was performed after adding the six cell type proportions as covariates to the statistical model.

5. EWAS of maternal ppBMI and placental DNAm

Robust linear regressions from the MASS R package[234] were run to account for potential heteroskedasticity and to test the associations between normalized placental DNAm beta values at each CpG site and maternal ppBMI as a continuous variable. Models were adjusted for maternal age, parity, maternal education and maternal smoking during pregnancy. Two EWAS models were run with and without adjustment for RefFreeCellMix putative cellular components.

6. Meta-analysis of the maternal ppBMI and placental DNAm EWAS

Cohorts that were members of the PACE consortium at that moment, had DNAm data from placental tissue obtained with the Illumina 450K or the EPIC BeadChips, and had maternal BMI information before pregnancy were invited to participate in the present study. The ten cohorts that contributed to the meta-analysis were Asking Questions about Alcohol in pregnancy (AQUA)[235], Early Autism Risk Longitudinal Investigation (EARLI)[236], EDEN[237], Genetics of Glucose regulation in Gestation and Growth (Gen3G)[238], Genetics, Early Life Environmental Exposures and Infant Development in Andalusia (GENEIDA)[239], Harvard Epigenetic Birth cohort (HEBC)[240], INMA[217], the InTraUterine sampling in early pregnancy study (ITU)[241], Markers of Autism Risk in Babies Learning Early Signs (MARBLES)[242], New Hampshire Birth Cohort Study (NHBCS)[243], and Rhode Island Child Health Study (RICHS)[244]. All cohorts had obtained ethics approval and informed consent from participants prior to data collection through their Institutional Ethics Boards. Exclusion criteria for the study were: non-singleton births, preeclampsia, and DNAm data not derived from the fetal-facing side of the placenta. All participants included in this meta-analysis were of European ancestry, and the DNAm quality control protocol was the same as the one described above for the INMA cohort. Since ITU and Gen3G cohorts analyzed their methylation data using EPIC arrays, only the probes also present in the 450K array were considered for the meta-analysis.

We performed an inverse variance-weighted, fixed effects meta-analysis using Genome-Wide Association Meta Analysis (GWAMA) software[245] with DNAm data from 419,460 CpG sites obtained with either the Illumina 450K or EPIC BeadChips in 2,631 placental samples of the fetal side. The meta-analysis was performed independently by two groups to ensure that consistent results could be reproduced. Results were fully consistent. We used the Bonferroni adjustment to correct for multiple testing. Secondary analyses were only performed on CpG sites that passed the Bonferroni

correction, particularly in the RefFreeCellMix-adjusted model. It is worth mentioning that both the meta-analysis and shadow meta-analysis teams performed an overall quality control separately prior to the meta-analysis itself, showing consistent results. In summary, we checked that conflictive probes had been removed and drew the cohort-specific Q-Q plots, as well as the correlation between sample sizes and significant hits across cohorts. Finally, we checked the genomic inflation of the whole meta-analysis and plotted forest plots of the significant hits after leaving one cohort out at a time, to see whether any of the cohorts was guiding the associations.

7. Functional and regulatory enrichment analyses of the maternal ppBMI-sensitive CpG sites

We annotated CpG sites to their closest genes and to Relation to CpG Island annotations from the FDb.InfiniumMethylation.hg19[246] and IlluminaHumanMethylation450kanno.ilmn12.hg19[247] R packages, as well as with several regulatory features using publicly available data: placental 15-chromatin states[248] released from the ROADMAP Epigenomics Mapping Consortium (ChromHMM v1.10)[249], placental germline DMP[250] and placental PMDs[40].

Over-representation analyses for gene sets or pathways were performed at the gene level with ConsensusPathDB[251] using Kyoto Encyclopedia of Genes and Genomes (KEGG)[252], Reactome[253], Wikipathways[254] and Biocarta gene sets. ConsensusPathDB performs a hypergeometric test with a default background equal to the number of ConsensusPathDB entities that are annotated with an ID of the type the user has provided and participate in at least one pathway[251]. Finally, the program corrects multiple testing with False Discovery Rate (FDR). Enrichment for TF was assessed at the gene level with EnrichR using the Encyclopedia of DNA Elements (ENCODE) and ChEA consensus TF from ChIP-X database[255]. EnrichR results were ranked using the combined score (P value computed using Fisher's exact test combined with the Z score of the deviation from the expected rank)[256].

8. Overlap of maternal ppBMI-sensitive CpG sites and birth outcomes associated SNPs

We assessed the genomic proximity between CpG sites identified in our maternal ppBMI placental EWAS (Bonferroni significant in the cell-type adjusted model) and SNPs previously associated with BW, birth length, head circumference and GA[40,248,249,257-259]. Briefly, we verified the genomic proximity between SNPs from the largest GWAS performed at that moment when analyses were performed on the abovementioned birth outcomes and our identified CpG sites by using the GenomicRanges package[260] in R, within 1Mb windows (±0.5Mb) surrounding each of the 367 autosomal SNPs.

9. Comparison of maternal ppBMI-sensitive CpG sites in placenta and in cord blood

We examined whether maternal ppBMI-associated CpG sites in placenta were the same as those previously reported in cord blood[57]. As no overlap was found between the hits that passed the Bonferroni correction in each study, we searched for CpG sites from the cord blood study present 0.5Mb upstream or downstream of each of the maternal ppBMI-associated CpG sites in placenta (1Mb windows) using the GenomicRanges R package[260], with the aim of finding genomic regions where DNAm was related to ppBMI in the two different tissues.

10. Placental *cis*-mQTL analysis

We performed the *cis*-mQTL mapping with paired genotype and methylation data in 368 samples with 4,171,035 SNPs and 747,486 CpG sites. The nominal modality of the TensorQTL[261] software was used to perform linear regressions between the genotype and the normalized DNAm RNT values, as implemented in FastQTL[262]. The covariates included in the regression model were the sex of the fetuses, the first five PCs derived from the genotype data (genotype PCs), the first 18 non-genetic DNAm PCs, and the cell type proportions calculated with the Planet methylation panel. Genotype and DNAm PCs were included in the model as covariates to remove hidden and/or technical confounders affecting the DNAm data. To avoid multicollinearity between the non-genetic DNAm PCs and the other covariates, the DNAm PCs were calculated on the residuals from a multiple linear regression adjusting the normalized DNAm RNT-values by the known covariates (sex of the fetuses, the first five genotype PCs and the five estimated cell types). Following Min *et al*. 2021, we kept all DNAm PCs that cumulatively explained 80% of the variance with a maximum number of 20 PCs for subsequent steps[263]. Then we performed a GWAS on each of the DNAm PCs and retained those PCs that were not associated with the genotype at a suggestive threshold ($P > 1 \times 10^{-7}$). This procedure returned 18 non-genetic DNAm PCs.

The TensorQTL *cis*-region was specified as ±0.5 Mb from each tested CpG site position, consistent with the results of previous studies where the distance between the majority of the *cis*-SNPs (mVariant) and the CpG sites (mSite) was ≤ 0.5 Mb[84,264-266]. In the final nominal *cis*-mQTL database we included only those probes with at least one *cis*-mQTL at $P_{nominal} < 5x10^{-8}$. This same regression model was used to build up two additional *cis*-mQTL databases; permuted and conditional. The permuted *cis*-mQTL database consists in correcting for multiple correlated variants tested via a beta approximation which models the permutation outcome with a beta distribution as described in Ongen *et al*. 2016. This database was presented as the main database and characterized accordingly, but the nominal database was used for the downstream

analyses, as suggested by the developers of the software and packages used. Finally, the conditional database uses the permuted QTLs to perform a stepwise regression procedure to map conditionally independent *cis*-QTLs as described by the GTEx Consortium[267].

Glossary

Along this dissertation three different *cis*-mQTL databases and two different and P values are mentioned in relation to TensorQTL, their description and nomenclature is described as follows:

- Nominal *cis*-mQTL database: it contains millions of association tests of all possible CpG site-SNP pairs in *cis* (i.e. SNPs located within a 0.5 Mb window, upstream and downstream each CpG position).
- $P_{nominal}$: P value obtained from the association between a CpG site and an SNP in the nominal database.
- Permuted *cis*-mQTL database: it corrects nominal *cis*-mQTL database for the multiple-correlated SNPs tested via a permutation scheme based on a beta approximation distribution of most significant $P_{nominal}$ (i.e. SNP) for a CpG site.
- P_{beta}: P value obtained from the association between a CpG site and an SNP in the permuted database.
- Conditional *cis*-mQTL database: it uses permuted *cis*-mQTL database to discover multiple independent SNPs (signals) per CpG site.
- $P_{conditional}$: P value obtained from the conditional analysis.

11. Placental *cis*-interacting mQTL analysis

The cell type-interacting mQTLs (imQTL) were computed with the interaction modality in TensorQTL, consisting in nominal associations for a linear model that includes a genotype per interaction term[261]. Additionally, we used the run_eigenmt = True option, to compute eigenMT-adjusted P values. This method runs faster than permutations, calculates adjusted P values that closely approximate empirical ones[268] and thus, has been implemented by TensorQTL as an alternative to permutation in imQTL mapping. The same SNPs, CpG sites and samples from the *cis*-mQTL analysis were considered for the interaction analyses. The covariates used were the sex of the fetuses, the first five genotype PCs and 18 DNAm PCs. Following the recommendations by Kim-Hellmuth *et al*. 2021, estimations of STB and TB proportions per sample were defined separately as the interaction terms. We considered significant only those imQTLs that showed an eigenMT FDR < 0.05.

We estimated sharing between STB- and TB-imQTLs with Storey's pi1 method[269]. Briefly, we took the eigenMT FDR < 0.05 STB-imQTLs and retrieved the nominal P values of these SNP-CpG pairs from the TB-imQTL set. Then we used the qvalue package[270] to estimate pi1 (the proportion of true positives). As STB and TB proportions are negatively correlated in term placentas, we wanted to know whether the interaction effect sizes of the overlapping imQTLs were also negatively correlated. Therefore, we retrieved the unique interacting mSites (imSite) for each one of the two cell types analyzed, as well as for GA, with a $P_{nominal} < 5x10^{-8}$ and drew a Venn diagram. Finally, with this list of imSites, we calculated the correlation of the interaction effect sizes of the overlapping imQTLs considering only the top interacting mVariant (imVariant) from the TB-imQTLs database.

Additionally, both GA and sex were also considered as interaction terms. In the first case, the covariates included were the sex of the fetuses, the first five genotype PCs, the 18 DNAm PCs and the cell type proportions from Planet. In the second case, the covariates included were the first five genotype PCs the 18 DNAm PCs and the cell

type proportions from Planet. For sharing calculations, the same methods as with TB were used for GA.

12. Functional and regulatory enrichment of the placental *cis*-mQTLs

As previously mentioned, the characterization of *cis*-mQTLs was performed on the permuted database in different steps. First, we plotted the distribution of both the distance between each mSite and mVariant pair, and the mQTLs along the chromosomes, followed by a volcano plot to check the distribution of the negative and positive effect sizes.

Second, using the annotation from the IlluminaHumanMethylationEPICanno.ilm10b4.hg19 R package[271], we performed several Chi-squared tests to assess the enrichment and depletion of our mSites in the University of California Santa Cruz (UCSC) RefGene and Relation to CpG Island annotations. We also depicted density plots of the DNAm values according to the Relation to CpG Island annotation. Moreover, for the top 10,000 mSites, we assessed enrichment and depletion of the overlap with tissue-specific and cell type-specific regulatory features, including DNase I hypersensitivity sites (DHS), all 15-state chromatin marks, and all five H3 histone marks (i.e., H3K27me3, H3K4me1, H3K4me3, H3K36me3, H3K9me3) from consolidated ROADMAP Epigenomics Mapping Consortium[249] using the experimentally-derived Functional element Overlap analysis of ReGions from EWAS (eFORGE) v2.0[272-274]. The enrichment and depletion with each of the three putative functional elements were tested separately and compared to the corresponding data from the consolidated ROADMAP epigenomics reference panel. Briefly, we selected only the top 10,000 probes following the advice of the developers of eFORGE to avoid saturating the background bins, especially those related to CpG Island annotation categories that may be more limited, taking into consideration the total amount of probes per category in the Illumina EPIC array.

Lastly, over-representation analyses of Gene Ontology (GO) and KEGG gene sets were conducted with MissMethyl R package[275-278]. MissMethyl performs a hypergeometric test considering the bias derived from the differing number of probes per gene and the multiple genes annotated per CpG site[278]. Gene set enrichment analyses with the Disease Ontology (DO) database were conducted with the DOSE R package[279]. With DOSE, we performed the gene set enrichment analysis considering the genes annotated in the IllumuminaHumanMethylationEPICanno.ilm10b4.hg19 R package[271] as background.

13. GWAS of neuropsychiatric traits

The GWAS summary-statistics used in this analysis were the largest, publicly available association studies of ADHD[280], AGR[281], ASD[282], BIP[283], INT[284], MDD[285], OCD[286], PD[287], SA[288], and SCZ[289] from several consortia and repositories, including Early Genetics and Lifecourse Epidemiology (EAGLE), Indonesia Schizophrenia Consortium, International Obsessive Compulsive Disorder Foundation Genetics Collaborative, OCD Collaborative Genetics Association Studies, PsychENCODE, Psychiatric Genomics Consortium (PGC), Psychosis Endophenotypes International Consortium, UK Biobank, SynGO Consortium, and 23andMe.

GWAS summary-statistics were harmonized to dbSNP build 155 with the INMA genotype data as a reference. The harmonization steps included: changing rsID to chromosome:position nomenclature, correcting the effect allele, the effect size and the allele frequency according to the reference genotypes if applicable, and creating a .ma format file with the summary-statistics as indicated in the SMR pipeline[121]. All locusZoom plots were constructed as suggested in their web page, using the LD panel of 1000 Genomes project (1000G)[290] as a reference.

14. Multi-SNP-based MR analysis between placental *cis*-mQTLs and neuropsychiatric disorders

MR analyses were carried out considering *cis*-mQTL SNPs as the IVs, CpG methylation as the exposure (X), and the neuropsychiatric traits as the outcome (Y). Multi-SNP-based MR (SMR-multi)[291] analysis was performed with the SMR software[121]. As depicted in the *Introduction* section, SMR integrates GWAS and QTL summary-statistics to test for pleiotropic associations between quantitative traits, such as DNAm or expression, and a complex trait of interest, including a disease. More precisely, SMR-multi includes multiple SNPs at a *cis*-mQTL locus in the SMR test to calculate the causative effect of an exposure on an outcome (b_{xy}). First, SMR-multi selected all SNPs with $P_{nominal} < 5x10^{-8}$ in the *cis* region (±0.5 Mb of the CpG). Second, it removed SNPs in very strong LD with the top associated SNP (LD $r^2 > 0.9$). Then, the causative effect (b_{xy}) from the exposure on the outcome was estimated at each of the SNPs and combined in a single test using an approximate set-based test accounting for LD among SNPs[291]. Additionally, the HEIDI test was performed. HEIDI uses multiple SNPs in a *cis*-mQTL region to distinguish pleiotropy from linkage. As it is described in Zhu *et al*. 2016, under the hypothesis of pleiotropy, where DNAm and a trait share the same causal variant, the b_{xy} values calculated for any SNPs in LD with the causal variant are identical. Therefore, testing against the null hypothesis that there is a single causal variant is equivalent to testing whether there is heterogeneity in the b_{xy} values estimated for the SNPs in the *cis*-mQTL region. For each DNAm probe that passed the genome-wide significance ($P_{nominal} < 5x10^{-8}$) threshold for the SMR test, HEIDI tested the heterogeneity in the b_{xy} values estimated for multiple SNPs in the *cis*-mQTL region. In this analysis, significant pleiotropic associations between DNAm and the neuropsychiatric diseases were selected as those with P_{SMR} corrected with Bonferroni $P < 0.05$ and $P_{HEIDI} > 0.05$ (not showing heterogeneity).

In the two interaction *cis*-mQTL databases used for SMR, we included probes with at least one *cis*-imQTL at $P_{nominal} < 5x10^{-8}$. The imQTLs obtained from each model were categorized as positively and negatively correlated with cell type estimates, or as uncertain if this was ambiguous, following Kim Hellmuth *et al*. 2021. To categorize the imQTLs into these groups, genotype main effects at low (< 25th percentile) vs high (> 75th percentile) cell type proportion were compared. imQTLs with positive cell type correlation showed an increase of the main genotype effect from low to high cell proportions (β_{mQTL} low < β_{mQTL} high). imQTLs with negative cell type correlation showed a decrease (β_{mQTL} low > β_{mQTL} high) and the uncertain group contained imQTLs for which sign flipped between low and high cell type proportions (β_{mQTL} low ≠ β_{mQTL} high).

15. Colocalization analysis between placental *cis*-mQTLs and neuropsychiatric disorders

Recalling the definition of *cis*-mQTLs in our analysis, all mSites with FDR < 0.05 and located within ±0.5 Mb of the regions defined in the original neuropsychiatric GWAS were tested. Colocalization analysis was performed to identify the overlap between causal mQTLs and GWAS-significant loci, as described in Giambartolomei *et al*. 2014, with the R Coloc package[113]. Only those neuropsychiatric traits with more than one hit in the SMR analysis were considered for the colocalization test. In total 12,228; 10,343; and 38,412 mSites were tested for BIP (n = 63 autosomal regions), MDD (n = 98 autosomal regions) and SCZ (n = 279 autosomal regions), respectively. Because data for all SNPs (regardless of significance) are required for this analysis, first, the mQTL analysis was rerun for these mSites so that association statistics could be recorded for all mVariants within ±0.5 Mb of the DNAm site. Then, we retrieved the regression coefficients, their variances and the SNP MAFs from the original GWAS data and our mQTLs, and the posterior probabilities (PP) were left as their default values. This methodology quantifies the support across the results of each GWAS for five hypotheses by calculating the PP, denoted as PPi for hypothesis H_i.

H_0: there exist no causal variants for either trait;

H_1: there exists a causal variant for one trait only, GWAS;

H_2: there exists a causal variant for one trait only, DNAm;

H_3: there exist two distinct causal variants, one for each trait;

H_4: there exists a single causal variant common to both traits.

16. Conditional analysis between placental *cis*-mQTLs and neuropsychiatric disorders

We performed a conditional analysis with the Genome-wide Complex Trait Analysis (GCTA) software[292,293] conditioning on the top *cis*-mQTL of those probes that passed SMR (Bonferroni $P_{SMR} < 0.05$) but failed to pass the HEIDI test ($P_{HEIDI} < 0.05$), due to the fact that heterogeneity may sometimes be driven by real secondary signals. We also performed the conditional analysis using GWAS summary data of the same set of SNPs (SNPs in the *cis*-mQTL region conditioning on the top *cis*-mQTL) for each of the three phenotypes that were studied in more detail (BIP, MDD, SCZ). For any of the regions showing evidence of a secondary signal ($P_{conditional} < 5x10^{-8}$) in either mQTL or GWAS data, we reran the conditional analyses in both mQTL and GWAS data conditioning on the secondary signal, and then used the conditional results for SMR and HEIDI tests. In this step, significant secondary pleiotropic associations between DNAm and the neuropsychiatric diseases were selected as those with Bonferroni $P_{SMR} < 0.05$ and $P_{HEIDI} > 0.05$ (not showing heterogeneity).

17. RICHS placental eQTMs

RICHS recruited mother and infant pairs from March 2009 to May 2013 following delivery at the Women and Infants Hospital of Rhode Island, United States. RICHS selected infants both small for GA, large for GA and controls born appropriate for GA matched according to sex, GA (±3 days), and maternal age (±2 years). The study protocol was approved by the Institutional Review Boards of Brown University and Women and Infants Hospital of Rhode Island. Placental RNA-seq data from a subset of samples (n = 200) were obtained using the Illumina Hi-Seq 2500 platform, aligned to the human reference genome, and RNA transcript abundance was quantified using Salmon[294]. About 20 million single-end RNA-seq reads were generated for each sample[295]. Placental DNAm data (n = 220) were obtained using the Illumina Infinium HumanMethylation450 BeadChip, preprocessed, and normalized with the minfi R package[230]. eQTMs were calculated by implementing linear models in MatrixEQTL, considering *cis*-windows of 0.5 Mb up and downstream of each CpG site in a total of 195 placenta samples. Linear regressions were adjusted for the sex of the fetuses, 5 PCs of expression and the Planet estimated cell types. Results were corrected with FDR.

The gene set enrichment analysis of the eQTM-genes associated to SCZ was performed using the Reactome pathway database and website analysis tool[296], with the default settings recommended by the developers in the Reactome database release 83 and the pathway browser version 3.7.

RESULTS

1. Maternal ppBMI is associated to 27 placental DNAm sites

Eleven North-American, Australian, and European studies (n = 2,631) contributed to the EWAS to determine the associations between maternal ppBMI and placental DNAm, including AQUA[235], EARLI[236], EDEN[237], Gen3G[238], GENEIDA[239], HEBC[240], INMA[217], ITU[241], NHBCS[243] and RICHS[244]. The effective sample size of the participating cohorts is depicted in Table 1. The MARBLES[242] cohort was excluded due to its small sample size and to the fact that forest plots showed inconsistencies with the rest of the cohorts.

Table 1. Effective sample size of the participating cohorts.

Cohort	Country	N	Array
AQUA	Australia	95	450K
EARLI	USA	54	450K
EDEN	France	664	450K
HEBC	USA	186	450K
Gen3G	Canada	448	EPIC
GENEIDA	Spain	103	450K
INMA	Spain	168	450K
ITU	Finland	352	EPIC
NHBCS	USA	311	450K
RICHS	USA	250	450K
TOTAL		2,631	

Maternal ppBMI (kg/m^2) was generally self-reported. In those cases where ppBMI was not available, BMI in early pregnancy (first trimester) was used. In all the analyses performed, we used ppBMI as a continuous variable. The cohort-specific average maternal ppBMI ranged from 22.9 in EDEN (France) to 27.6 in EARLI (United States)

(SD = 1.60). In general, 548 (20.8%) and 369 (14.0%) mothers reported overweight (BMI > 25) and obesity (BMI > 30), respectively, while 115 (4.37%) appeared to be underweight (BMI > 18.5). The distribution of the other covariates for each cohort is provided in Appendix 1.

Each cohort analyst conducted two different EWAS modelling DNAm beta values at a maximum number of 419,460 CpG sites in relation to maternal ppBMI using robust linear regressions, with and without adjustment for putative cellular components using RefFreeCellMixture[233]. Both models were adjusted for maternal age, parity, maternal education, and maternal smoking. Genomic inflation factors (λ) from the cohort-specific models ranged from 0.692 to 1.472 in NHBCS and EARLI, respectively. Finally, after quality control of the results, we conducted an inverse variance-weighted fixed-effects meta-analysis using the software GWAMA[245]. The inflation factors from the meta-analyses (λ = 1.230 and 1.220 for the cell-type adjusted and unadjusted models, respectively) revealed potential residual confounding and moderate inflation of test statistics.

After applying the Bonferroni correction for multiple-testing (meta-analysis P < 1.20×10^{-7}), we obtained 27 and 42 CpG sites at which maternal ppBMI was significantly associated with placental DNAm levels in the models adjusted (Figure 7) and unadjusted for cell type components, respectively. Full results for both models are provided in Appendix 2 and 3. Higher maternal ppBMI was associated with lower placental DNAm in 24/27 differentially methylated CpG sites identified in the cell type adjusted model, while in the unadjusted model, 33/42 hits showed positive associations (higher maternal ppBMI associated with increased placental DNAm at the identified CpG sites). However, effect sizes of CpG sites that were differentially methylated in one model were positively correlated to the effect sizes of the same position in the other. Finally, the heterogeneity of associations across cohorts was lower for the model adjusted for cell type proportions compared to the unadjusted model (26/27 vs. 34/42 CpG sites presented Cochran's Q-test P values > 0.01) and thus, we continued with the results from the fully adjusted model for all downstream analyses.

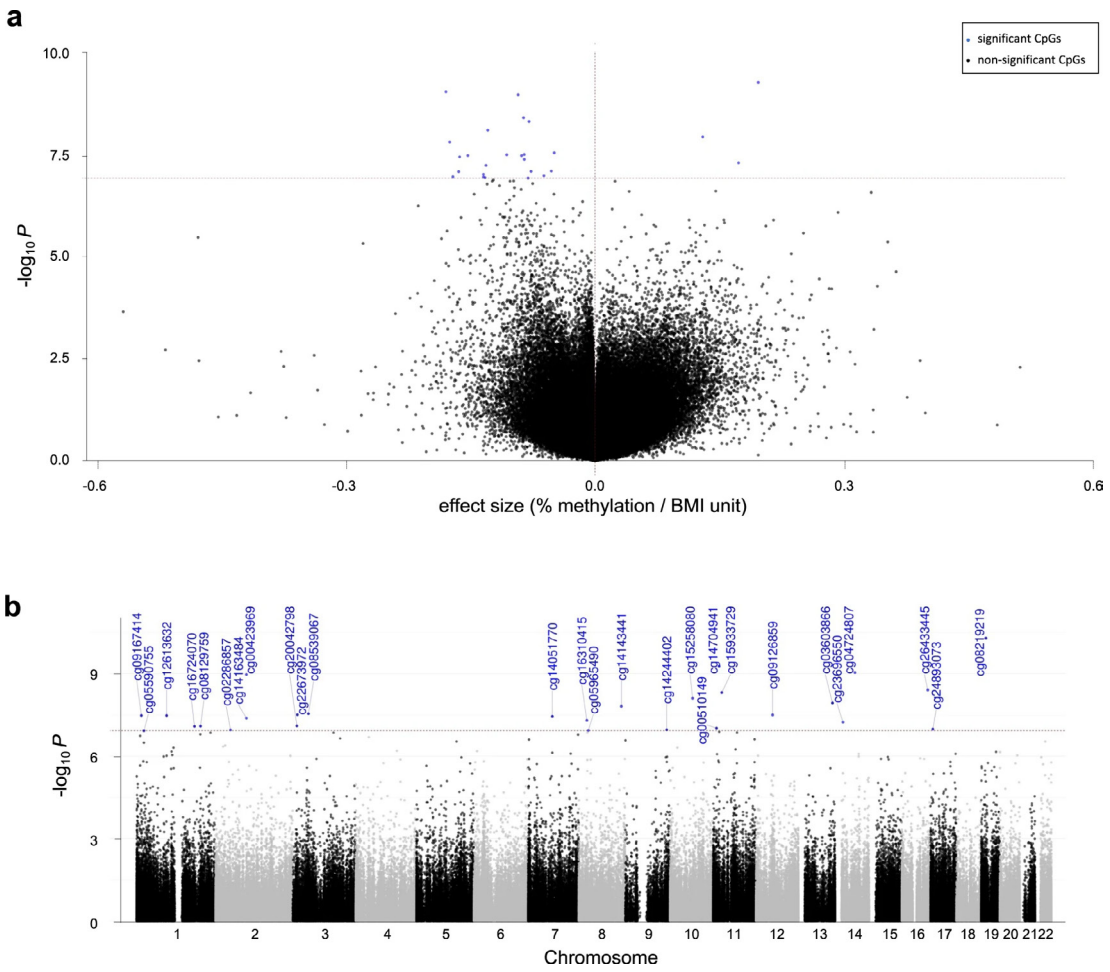


Figure 7. Association between maternal ppBMI and placental DNAm (n = 2,631 placental DNA samples), after adjusting for maternal age, parity, maternal education, maternal smoking and putative cellular heterogeneity.

Association results are displayed as **a** volcano plot, where the X-axis shows the effect sizes (ranging between 0 and 1) in DNAm and **b** Manhattan plot, where the X-axis represents the genomic location of each CpG. In both panels blue dots indicate significantly associated CpG sites (meta-analysis P < 1.20×10^{-7}).

Among the 27 DMPs identified in our cell type-adjusted EWAS (Figure 7), a few individual CpG sites are worthy of mention. The most notable association was observed at cg08219219, located in the eighth exon of *EPHX3*, with the lowest P value (Bonferroni P = 2.12×10^{-5}) and an effect size of -1.12×10^{-3}, meaning that a 10-unit difference in maternal ppBMI is associated with a 1.1% lower DNAm at this specific CpG site. This association was consistent across all cohorts (Cochran's Q-test P =

0.12) (Figure 8A). The largest positive effect size was observed in cg14704941, in the first intron of *CSRP3*, with an effect size of $1.96x10^{-3}$, corresponding to a 2% higher placental DNAm per 10-unit BMI (Bonferroni P = $2.21x10^{-4}$ and Cochran's Q-test P = 0.09) (Figure 8B). In turn, the largest negative effect size was found in cg04724807 (more than 57 kb upstream *SYT16*) with 1.8% lower DNAm per 10-unit ppBMI (Bonferroni P = $2.83x10^{-4}$, Cochran's Q-test P = 0.097) (Figure 8C). The following CpG sites reached the Bonferroni significance threshold and were not heterogenous across cohorts: cg00423969 and cg14163484, 1.5 kb upstream of the *FER1L5* promoter, as well as cg26433445, cg15933729 and cg08539067, close to *CMIP*, *LGR4* and *GPX1*, respectively (Figure 8D, E, F, G, H).

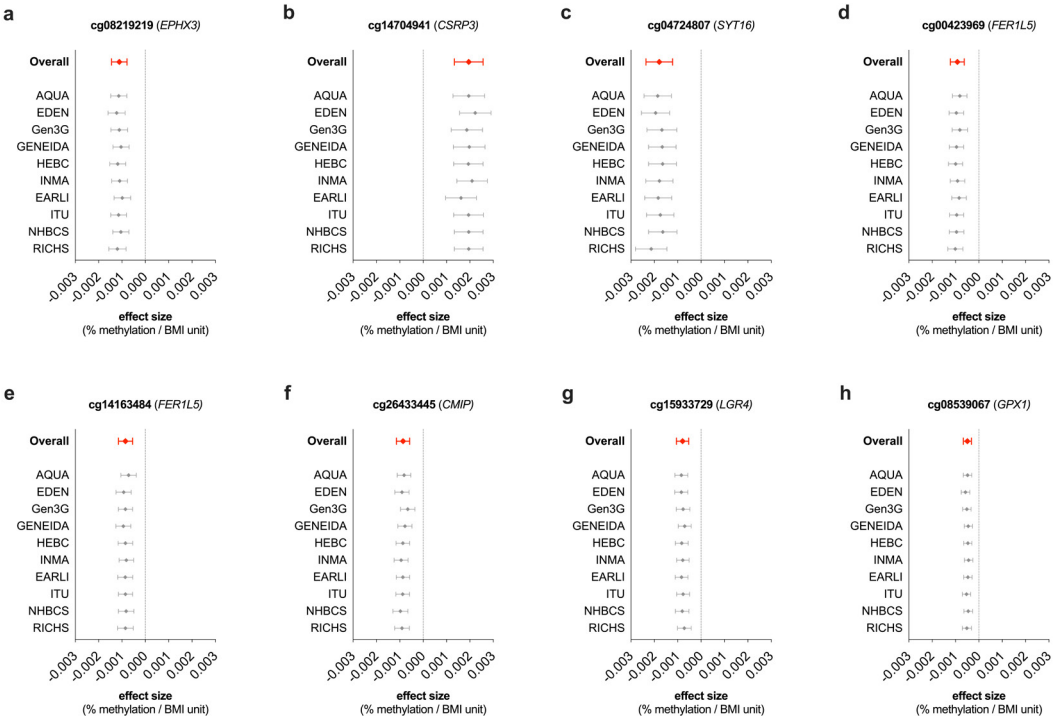


Figure 8. Forest plots of the leave-one-out analysis showing the fixed effects meta-analysis estimates of association between maternal ppBMI and placental DNAm.

Association of **a** cg08219219, **b** cg14704941, **c** cg04724807, **d** cg00423969, **e** cg14163484, **f** cg26433445, **g** cg15933729, and **h** cg08539067 with maternal ppBMI. In all panels, cohort names indicate the cohort excluded in each row, and error bars represent the 95% confidence interval of the effect size. Numerical source data for the figure are available in file Appendix 4.

To gain insight into the biological processes that may be captured by placental DNAm associated with maternal ppBMI, we performed gene set and regulatory enrichment analyses. To this end, first, we annotated CpG sites to genes and regulatory elements as explained in the *Methods* section. Then, we conducted gene set enrichments for the 26 unique genes annotated to the 27 maternal ppBMI-sensitive CpG sites with ConsensusPathDB[251] using KEGG[252], Reactome[253], Wikipathways[254] and Biocarta as reference databases. Two gene set pathways were significantly enriched (Q value < 0.05), namely small cell lung cancer and oxidative stress-induced signaling pathway. This was also true when we reduced the background from the default to only the genes that are represented in the Illumina 450K array (n ≈21,231). We also tested whether the genes annotated to maternal ppBMI-associated CpG sites were enriched for regulatory regions of specific TF with EnrichR[256] using ENCODE and ChEA consensus TF from the ChIP-X database[255]. Most notably, our ppBMI-associated CpG sites were enriched for genes regulated by *ZNF217* (P = 0.02).

2. The ppBMI-sensitive CpG sites are close to SNPs associated with BW

We wanted to determine whether the maternal ppBMI-associated CpG sites identified here were localized near genetic variants that have been associated with birth outcomes in previously published GWAS. Thus, we investigated whether our significant CpG sites were within ±0.5 Mb of SNPs that have been associated with BW (n = 310), birth length (n = 5), head circumference (n = 3), GA (n = 6), and BW + GA (n = 6)[40,248,249,257-259]. Of the 330 birth outcome SNPs in autosomal chromosomes, 10 BW-associated SNPs were within ±0.5 Mb of the CpG sites that were associated with maternal ppBMI. Therefore, more than a third of the 27 ppBMI-associated CpG sites were within ±0.5 Mb of BW SNPs, including cg00423969 (*FER1L5*), cg14163484 (*FER1L5*), cg00510149 (*IRAG1*), cg02286857 (*TTC7A*), cg15258080 (*HK1*), cg22673972 (*SLC6A6*) and cg24893073 (*KDM6B*).

3. The ppBMI-sensitive CpG sites are placenta-specific and not covered by cord blood studies

We assessed whether the DNAm signature of maternal ppBMI in placenta was consistent with associations in cord blood previously reported by the PACE consortium[57]. We did not find any overlapping CpG sites associated with maternal ppBMI between the two tissues. However, we reported three maternal ppBMI-associated CpG sites in the placenta that were less than 0.5 Mb upstream or downstream from CpG sites that had been associated to maternal ppBMI in cord blood: two out of three loci identified showed consistent effect directions of the association to maternal ppBMI in both tissues. For more details on the loci mentioned, see Appendix 5.

4. More than 200,000 placental CpG sites are regulated by genetic variants in *cis*

We discovered *cis*-mQTLs for 214,830 mSites with FDR < 0.05 (permuted mQTLs database) in 368 fetal placenta DNA samples from the Gipuzkoa, Sabadell and València cohorts of the INMA project[217]. Briefly, TensorQTL modelled the methylation RNT values from 747,486 CpG sites as a function of the genotype from 4,171,035 SNPs in a ±0.5 Mb window, and the model was adjusted for fetal sex, 5 genotype PCs, 18 DNAm PCs, and Planet[42]-estimated cell types, as covariates. Phenotypic information of donor mothers is summarized in Table 2. In the vast majority of the *cis*-mQTLs, the mVariant and the mSite were located relatively close to each other, with a median absolute distance of 13,87 kb, indicating that genetically modulated DNAm is typically close to the implicated regulatory variant, as has been previously described by other authors[263] (Figure 9A). The distribution of mQTLs across chromosomes tended to be in line with chromosome length (Figure 9B). Additionally, the HRC[222] reference alleles in placental *cis*-mQTLs presented both positive and negative effect directions (Figure 9C), as expected.

Table 2. Distribution of maternal smoking during pregnancy, demographic variables, birth outcomes and covariates of the INMA cohort.

Variable	Category	INMA (n = 368)		
		N	Mean ± SD or %	N missing
Maternal smoking during pregnancy	Yes	52	14.13%	5
	No	311	84.51%	
Maternal age (continuous)	Mean (SD)	368	30.85 ± 3.95	0
Parity	0	208	56.52%	0
	≥1	160	43.47%	
Maternal education	Primary or without education	68	18.47%	0
	Secondary	172	46.73%	
	University	128	34.78%	
BW (g)	Mean (SD)	368	3,278 ± 440.55	0
GA (weeks)	Mean (SD)	368	39.71 ± 1.30	0
Preterm birth (<37 weeks)	Yes	10	2.71%	0
	No	358	97.28%	
Ancestry	White	347	94.29%	5
	Non-white	16	4.34%	
Sex of child	Female	189	51.35%	0
	Male	179	48.64%	
Socioeconomic status	I+II	130	35.32%	0
	III	100	27.17%	
	IV+V	138	37.50%	

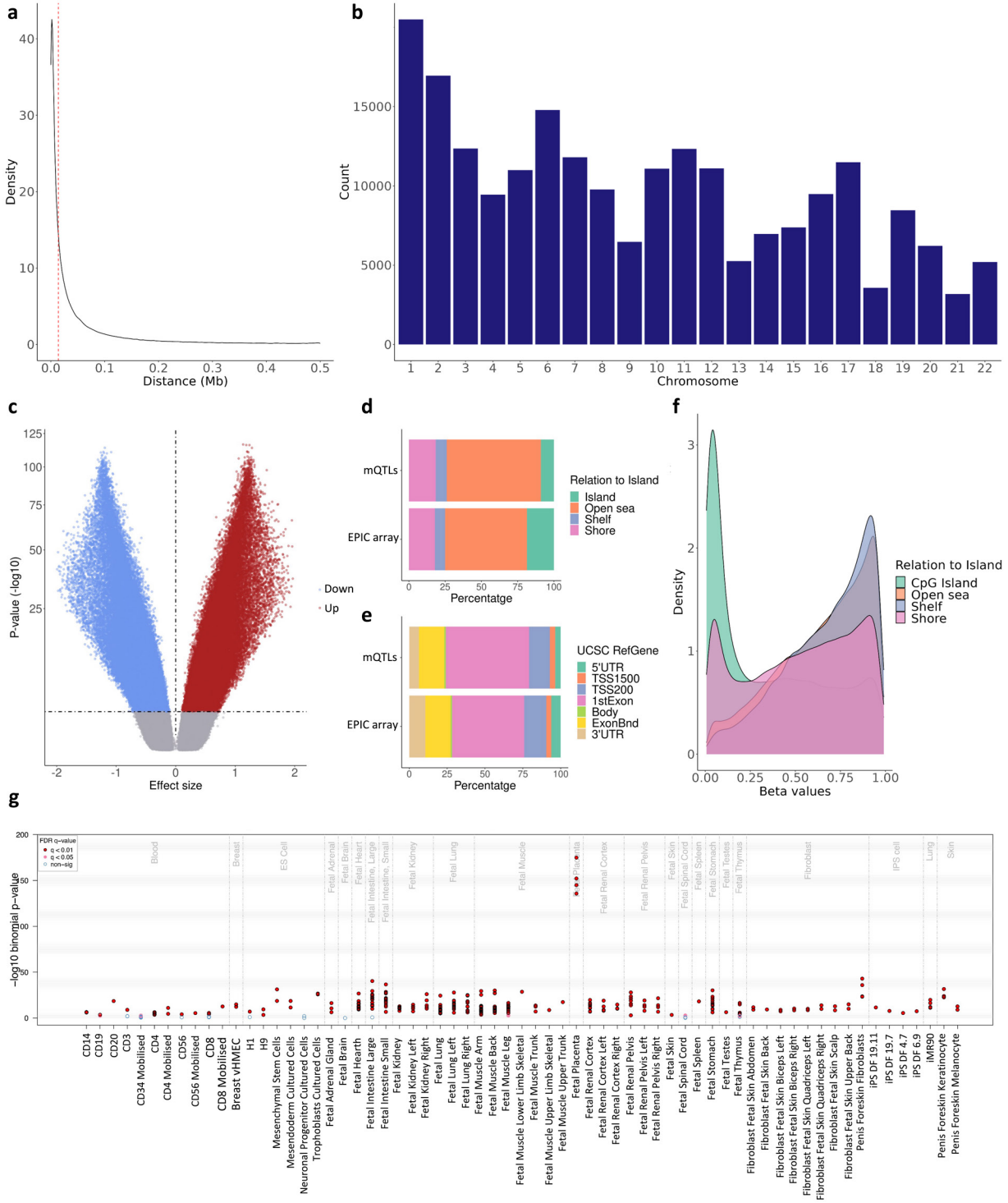
In general, mSites were depleted from CpG islands and promoters ($P < 2.2x10^{-16}$) and enriched within gene bodies ($P < 2.2x10^{-16}$) and genomic features with intermediate methylation values such as open sea, and CpG island shelf and shore regions ($P < 2.2x10^{-16}$) (Figure 9D, E and F). They were also enriched in the human leukocyte antigen (HLA) region with 2,896 out of the total 214,830 mSites of the permuted database ($P = 1.45x10^{-10}$). Using eFORGE[272-274], we were able to detect an enrichment in DHS

hotspots and H3K4me1 broadPeaks in different tissues, fetal placenta showing the strongest signal (Figure 9G). DHS hotspots mark accessible chromatin while H3K4me1 broadPeaks appear in enhancer regions. Afterwards, we annotated the mSites to their closest genes, performed a gene set enrichment analysis with the DO database[297], and obtained 219 enriched gene sets, including neuropathy (Benjamini-Hochberg P = $1.88x10^{-9}$), mood disorder (Benjamini-Hochberg P = $1.22x10^{-7}$), demyelinating disease (Benjamini-Hochberg P = $2.48x10^{-6}$), and BIP (Benjamini-Hochberg P = $6.96x10^{-6}$). Enrichment results can be found in Appendices 5, 6, and 7.

Apart from the permuted *cis*-mQTLs, in which correction for multiple correlated variants is performed via a beta approximation, nominal and conditional *cis*-mQTLs were also calculated in our 368 placenta DNA samples. Briefly, the nominal *cis*-mQTL approach reports all the SNP-CpG combinations while the conditional method uses the permuted QTLs to perform a stepwise regression procedure and map conditionally independent *cis*-QTLs. The three complete placental *cis*-mQTL databases are publicly available online in the following address: https://irlab.shinyapps.io/shiny_mqtl_placenta/.

Figure 9. Characterization of the placental *cis*-mQTLs from the permuted database. ▶

The distance between the SNP-CpG pairs participating in the reported *cis*-mQTLs is displayed as **a** density plot, where the X-axis represents the absolute distance in Mb. The red line represents the absolute median SNP-CpG distance. The distribution of the reported *cis*-mQTLs along the chromosomes is shown in the **b** barplot, where the X-axis represents the autosomal chromosomes. The distribution of the effect sizes from the reported *cis*-mQTLs is pictured in the **c** volcano plot, where the Y-axis illustrates the -log10 beta P value, and the X-axis the effect size. The blue and the red dots represent the mQTLs with a negative and positive effect, respectively. The distribution of the EPIC array probes and the nominal mSites considering the Relation To Island and the UCSC RefGene annotation is displayed in the **d** and **e** barplots, respectively. The DNAm beta values of the participating mSites stratified by the Relation To Island annotation is shown in the **f** density plot, where the DNAm values are found in the X-axis. The eFORGE enrichment of DHS hotspots considering the top 10,000 permuted mSites is shown in the **g** plot. The Y-axis represents the -log10 binomial P value of the enrichment, and the X-axis the tissues studied. FDR corrected Q values below 0.01 and 0.05 are represented by red and pink dots, respectively, while blue dots show Q values > 0.05.


a
Density
Distance (Mb)
b
Count
Chromosome
c
P-value (-log10)
Effect size
Down
Up
d
mQTLs
EPIC array
Percentatge
Relation to Island
Island
Open sea
Shelf
Shore
e
mQTLs
EPIC array
Percentatge
UCSC RefGene
5'UTR
TSS1500
TSS200
1stExon
Body
ExonBnd
3'UTR
f
Density
Beta values
Relation to Island
CpG Island
Open sea
Shelf
Shore
g
−log10 binomial p−value
FDR q-value
q < 0.01
q < 0.05
non−sig
Blood
Breast
ES Cell
Fetal Adrenal
Fetal Brain
Fetal Heart
Fetal Intestine, Large
Fetal Intestine, Small
Fetal Kidney
Fetal Lung
Fetal Muscle
Placenta
Fetal Renal Cortex
Fetal Renal Pelvis
Fetal Skin
Fetal Spinal Cord
Fetal Spleen
Fetal Stomach
Fetal Testes
Fetal Thymus
Fibroblast
iPS cell
Lung
Skin
CD14
CD19
CD20
CD3
CD34 Mobilised
CD4
CD4 Mobilised
CD56
CD56 Mobilised
CD8
CD8 Mobilised
Breast vHMEC
H1
H9
Mesenchymal Stem Cells
Mesendoderm Cultured Cells
Neuronal Progenitor Cultured Cells
Trophoblasts Cultured Cells
Fetal Adrenal Gland
Fetal Brain
Fetal Hearth
Fetal Intestine Large
Fetal Intestine Small
Fetal Kidney
Fetal Kidney Left
Fetal Kidney Right
Fetal Lung
Fetal Lung Left
Fetal Lung Right
Fetal Muscle Arm
Fetal Muscle Back
Fetal Muscle Leg
Fetal Muscle Lower Limb Skeletal
Fetal Muscle Trunk
Fetal Muscle Upper Limb Skeletal
Fetal Muscle Upper Trunk
Fetal Placenta
Fetal Renal Cortex
Fetal Renal Cortex Left
Fetal Renal Cortex Right
Fetal Renal Pelvis
Fetal Renal Pelvis Left
Fetal Renal Pelvis Right
Fetal Skin
Fetal Spinal Cord
Fetal Spleen
Fetal Stomach
Fetal Testes
Fetal Thymus
Fibroblast Fetal Skin Abdomen
Fibroblast Fetal Skin Back
Fibroblast Fetal Skin Biceps Left
Fibroblast Fetal Skin Biceps Right
Fibroblast Fetal Skin Quadriceps Left
Fibroblast Fetal Skin Quadriceps Right
Fibroblast Fetal Skin Scalp
Fibroblast Fetal Skin Upper Back
Fibroblasts
iPS DF 19.11
iPS DF 19.7
iPS DF 4.7
iPS DF 6.9
IMR90
Penis Foreskin Keratinocyte
Penis Foreskin Melanocyte

5. Cell type-interacting mQTLs are detected for STB and TB cells

The most abundant cells in fetal placenta are STBs. During gestation, undifferentiated TBs change into fully differentiated STBs, a continuous, multinucleated and specialized layer of epithelial cells[44]. We estimated the cell type proportions of our samples with the reference-based method Planet[42]. As expected, the estimated STB-content in our samples was negatively correlated with the estimated proportion of TBs ($P < 2.2x10^{-16}$, r = -0.89) and positively correlated with GA ($P < 4.4x10^{-5}$, r = 0.21) (Figure 10A and B). We used the interaction modality of TensorQTL with the eigenMT[268,298] correction to calculate STB-, TB- and GA-imQTLs, as proxies for STB-, TB- and GA-specific *cis*-mQTLs, and obtained 38 and 1 STB- and TB-imQTLs, respectively, with no significant GA-imQTL at a significance threshold of eigenMT FDR = 0.05. The higher amount of STB-imQTLs revealed a higher statistical power for the most abundant cell type, as previously pointed out by Kim-Hellmuth *et al.* The lists of STB- and TB-imQTLs can be found in Appendix 9.

The only significant TB-imQTL was also a STB-imQTL. We also estimated sharing between STB- and TB-imQTLs with Storey's pi1 method[269]. Briefly, we took the eigenMT FDR < 0.05 STB-imQTLs and calculated pi1 in the nominal P values of these SNP-CpG pairs in the unfiltered TB- and GA-imQTL sets. STB-TB sharing was estimated to be of 0.68, while STB-GA sharing resulted in a pi1 = 0.51. Beyond the global estimated sharing between cell types, and as STB and TB proportions are negatively correlated in term placentas, we wanted to know whether the genotype-cell type interaction had opposite effect directions in the overlapping imQTLs. Due to the limited number of eigenMT-significant imQTLs, we retrieved the unique imSites for each one of the two cell types analyzed, as well as for GA, with $P_{nominal} < 5x10^{-8}$. We found an overlap of 12 imSites between the two cell types, and no overlap between STB- and GA-imSites. For each of the cell type-overlapping imSites, we considered the top TB-imQTL, and we calculated the correlation between their interaction effect sizes and the effect

sizes reported for those same imQTLs in the STB-imQTL set. As a result, and despite the limited number of points in the regression, we obtained a highly significant negative correlation ($P = 4.1x10^{-10}$ and $r = -0.99$), as in the example in Figures 10D, E and F. Of note, the imQTL in Figure 10 is the only eigenMT-significant TB-imQTL (FDR = 0.05). The moderate sharing between STB- and GA-imQTLs revealed by Storey's pi1 suggests that the effect of GA on the genetic regulation of placental DNAm is at least partially independent of the TB-to-STB transition. Finally, as placental DNAm could differ between sexes, we also mapped sex-imQTLs but obtained no significant imQTLs after eigenMT FDR adjustment.

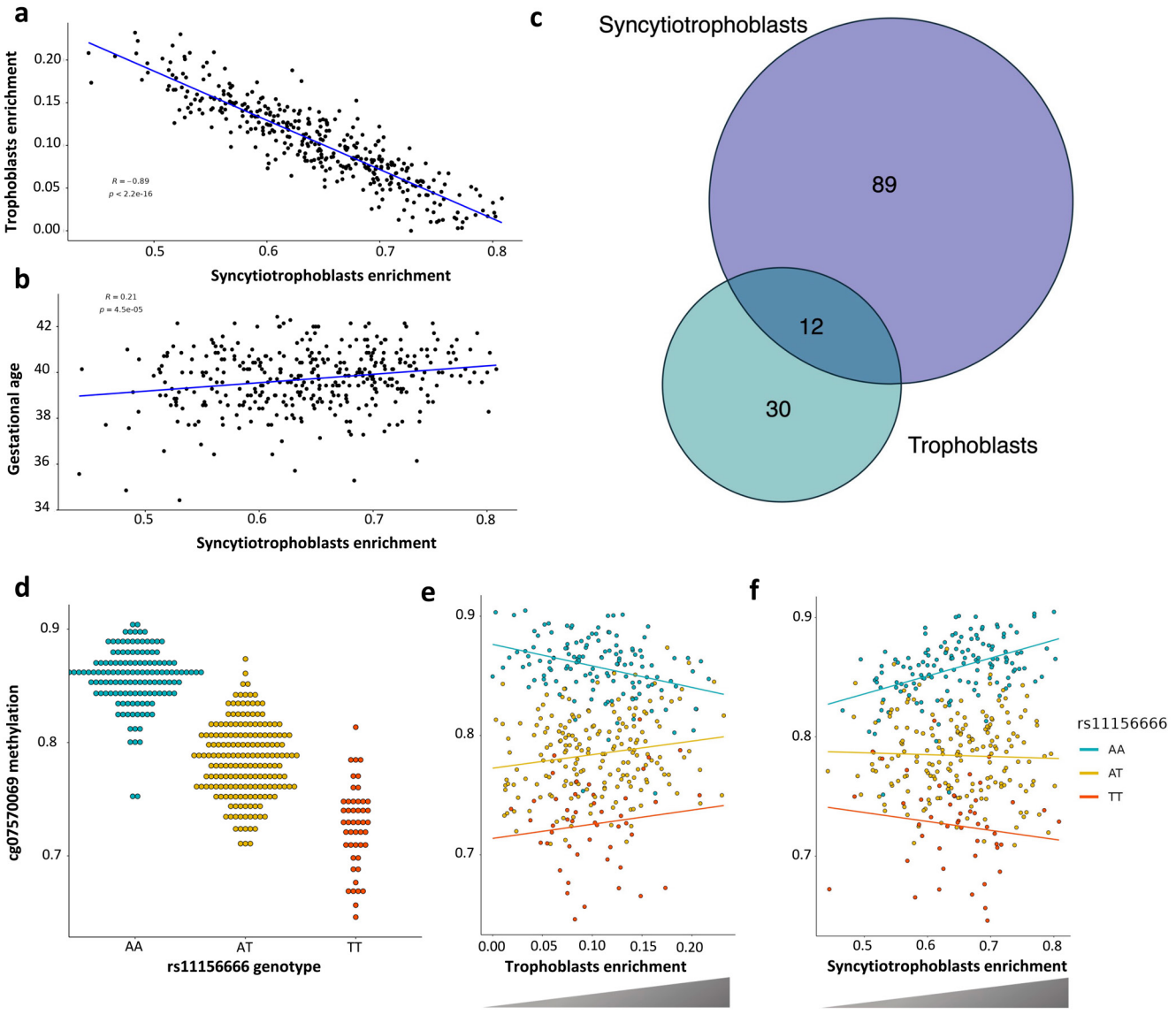


Figure 10. STB- and TB-imQTLs.

The correlation between STB and TB cell types (n = 368) is shown in **a**, the X-axis representing the STB proportion in the sample set, and the Y-axis showing the estimated TB proportion. The correlation between STB and GA (n = 368) is shown in **b**, the X-axis representing the STB proportion in the sample set, and the Y-axis showing the GA. The intersection between STB- and TB-imQTLs is represented in the **c** Venn diagram. The standard cg07570069-rs7150262 mQTL, as well as the TB- and STB-imQTLs, are displayed in the **d**, **e** and **f** dotplots, respectively. In all three, the Y-axes represent the cg07570069 DNAm beta values, ranging from 0 to 1. In **d**, the X-axis displays the genotype of rs7150262, while in **e** and **f**, the X-axes show the TB and STB proportions, respectively. In the case of **e** and **f**, the genotype of rs7150262 is color coded. The P values of the (i)mQTLs were the following: $P_{nominal} = 1{,}365x10^{-97}$, eigenMT FDR = $2.362x10^{-2}$, and eigenMT FDR = $1.579x10^{-4}$ for the standard mQTL, and the TB- and STB-imQTLs, respectively.

6. Multi-omics approaches unravel the potential placental origin of some neuropsychiatric disorders

We selected the largest available GWAS for each of the following traits: ADHD[280], AGR[281], ASD[282], BIP[283], INT[284], MDD[285], OCD[286], PD[287], SA[288] and SCZ[289] (Table 3). Most of the studies were carried out in the context of either the PGC[299] or the EAGLE[300] consortia. GWAS sample sizes ranged from 6,183 in the case of SA and 500,199 samples in MDD (Table 3).

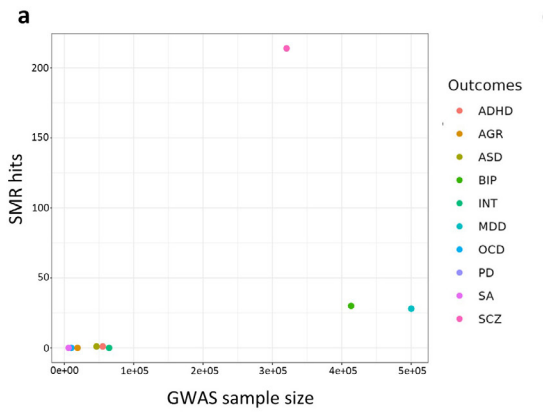
Table 3. Overview of the neuropsychiatric traits and disorders included in this study.

Trait	Abbreviation	N	N cases	N controls	Original N of SNPs	GWAS significant loci	N SMR hits	Ref.
Attention-deficit and hyperactivity disorder	ADHD	55,374	20,183	35,191	8,094,094	12	1	280
Aggression	AGR	18,988			2,188,528	1	0	281
Autism spectrum disorder	ASD	46,351	18,382	27,969	9,112,386	3	1	282
Bipolar disorder	BIP	413,466	41,917	371,549	7,608,183	64	30	283
Internalizing problems	INT	64,561			5,445,594	0	0	284
Major depression disorder	MDD	500,199	170,756	329,443	8,483,301	101	28	285
Obsessive-compulsive disorder	OCD	9,725	2,688	7,037	8,409,517	0	0	286
Panic disorder	PD	10,240	2,248	7,992	10,151,624	0	0	287
Suicidal attempt	SA	6,183	3,288	2,895	11,823,118	0	0	288
Schizophrenia	SCZ	320,404	76,755	243,649	7,583,660	287	214	289

We searched for placenta DNAm sites pleiotropically associated with neuropsychiatric disorders using the SMR and HEIDI tests included in the SMR software[121]. We used the summary statistics of the abovementioned GWAS and the nominal placental *cis*-mQTL database filtered by $P_{nominal} < 5x10^{-8}$, as suggested by the developers of SMR. Importantly, all 130,736 mSites considered for SMR were FDR significant in the permuted database. No significant hits were found for many of the traits, such as OCD or SA. In turn, SMR hits (Bonferroni $P_{SMR} < 0.05$ and $P_{HEIDI} > 0.05$) were identified for ADHD (n = 1), ASD (n = 1), BIP (n = 30), MDD (n = 28) and particularly SCZ (n = 214). The complete SMR results can be found in Appendix 10. Five hits were common to BIP and SCZ, with the same effect direction. Additionally, most of those hits did not overlap with CpG sites reported to change as a function of GA[301], suggesting there is a relative stability in the DNAm levels during pregnancy.

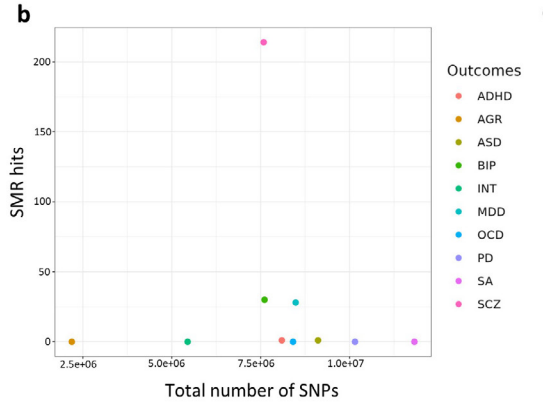
However, we wanted to make sure that our SMR results were not merely led by statistical power. Therefore, we calculated the correlation between the SMR hits and i) the GWAS sample size, ii) the total number of SNPs per GWAS, and iii) the number of GWAS-significant loci, in all the neuropsychiatric disorders studied. We observed a correlation of r = 0.48 between the number of significant SMR results and the GWAS sample size, although it did not reach statistical significance, probably due to the limited number of SMR experiments performed (Figure 11A). The total number of SNPs included in each of the GWAS was not correlated at all with the results obtained in SMR, but there was a highly significant correlation between the latter and the number of significant loci reported in each GWAS ($P < 1.7x10^{-6}$, r = 0.97) (Table 3, Figure 11B and C). Anyway, it is worthy of mention that SCZ is by far the disorder showing the highest ratio between SMR hits and associated GWAS loci, suggesting that placental methylome may be particularly relevant for this disorder. Additionally, we ran the same SMR approach with two previous, smaller SCZ GWAS and still obtained significant SMR hits, with a remarkable overlap among the three studies. The characteristics of those studies are summarized in Figure 11D, together with the Venn diagrams showing the overlapping hits (Figures 11E and F). In summary, our results reveal an at least moderate correlation between the number of SMR hits and the GWAS sample size, that is surpassed by far by the number of GWAS significant loci, with a large effect on the capacity of SMR to find potentially interesting results.

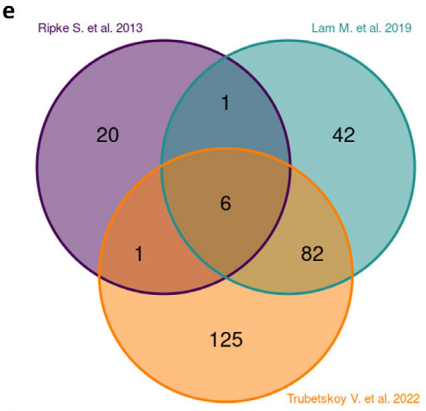
Nevertheless, the abundant results in SCZ are not fully explained by this effect.

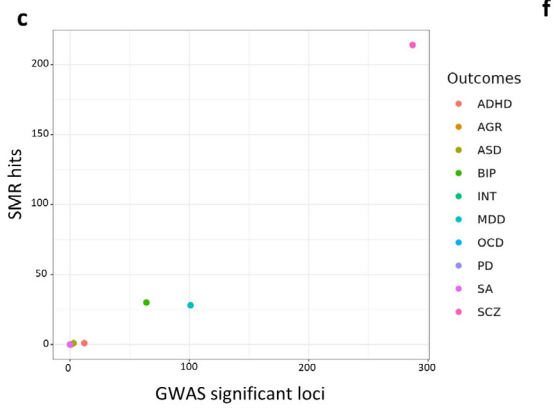


d

Author	Ripke S. et al.2013	Lam M. et al.2019	Trubetskoy V. et al.2022
PubMed ID	23974872	31740837	35396580
N of cases	5,001	56,418	76,755
N of controls	6,243	78,818	243,649
N of total samples	11,244	135,236	320,404
N of SMR hits	52	131	214







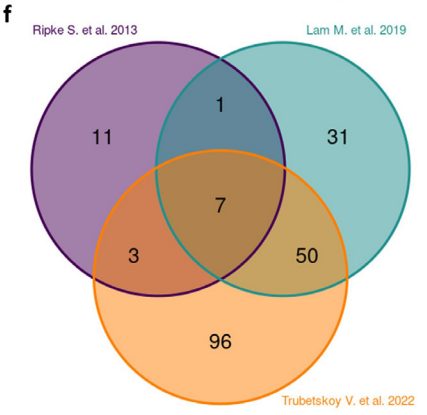


Figure 11. Correlation between the statistical power of GWAS and SMR results.

We calculated the correlation between the number of SMR hits and **a** the GWAS sample size, **b** the total number of SNPs per GWAS, and **c** the GWAS significant loci in all the neuropsychiatric disorders studied. We also run SMR with two previous SCZ GWAS (apart from Trubetskoy *et al.* 2022. The characteristics of these studies are summarized in table **d**, and the overlap among studies in terms of **e** mQTLs and **f** mSites are represented as Venn diagrams.

To replicate the pleiotropic associations identified in the SMR analyses, we performed colocalization tests with the Coloc R package[113]. Colocalization was performed across the 64, 102, and 287 genomic regions defined in the GWAS associated with BIP, MDD and SCZ, respectively. Regarding mQTLs, from the total amount of 214,830 mSites with FDR < 0.05 in the permuted database, 12,228, 10,343, and 38,412 mSites were considered for BIP, MDD and SCZ colocalization analyses, respectively. In BIP, the PP for 64 regions, involving 6,487 DNAm sites in 7,123 region-CpG pairs, were supportive of a colocalization signal (PP3+PP4 > 0.9) (Figure 12A). In MDD, the PP for 79 regions, involving 3,850 DNAm sites in 3,895 region-CpG pairs, were supportive of a colocalization signal (PP3 + PP4 > 0.9) (Figure 12B). Finally, in SCZ, the PP for 274 regions, involving 19,661 DNAm sites in 21,071 region-CpG pairs, were supportive of a colocalization signal (PP3+PP4 > 0.9) (Figure 12C). The Coloc results can be found in Appendix 11. When the overlap with SMR hits was assessed, out of the 30 SMR hits in BIP, 28 DNAm sites (93%) showed evidence of colocalization with BIP; in MDD, 20 DNAm sites (71%) out of the 28 SMR hits showed evidence of colocalization; and from the 214 SMR hits in SCZ, 198 DNAm sites (92%) colocalized with SCZ.

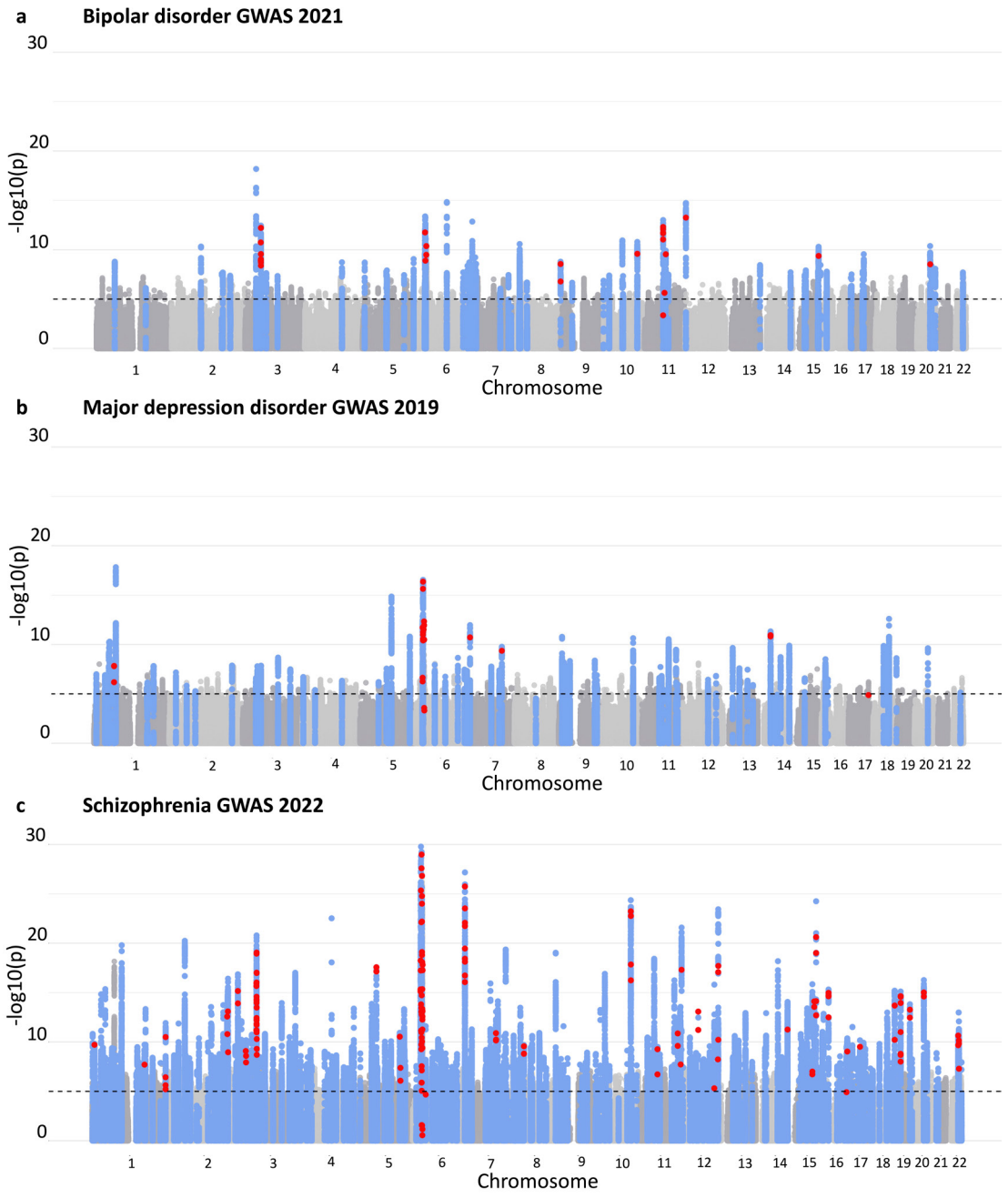


Figure 12. Manhattan plot of BIP, MDD and SCZ GWAS highlighting the SMR and Coloc results.

The original GWAS from BIP, MDD and SCZ were plotted in the **a**, **b** and **c** Manhattan plots, respectively. In the Y-axes the original -log10 P values are displayed, and in the X-axis the chromosomes. The blue dots represent genomic regions significantly colocalizing with our placental mQTLs, and the red dots are SNPs associated with CpG sites that have been shown to pleiotropically associate with either BIP, MDD or SCZ in the SMR approach. Therefore, the blue dots represent the colocalization results, and the red dots show the SMR results.

To provide more insights into the functional relevance of our findings, we examined the association between placental DNAm and placental gene expression in the CpG sites from the SMR and colocalization analyses. We interrogated eQTMs from an independent set of 195 fetal placenta samples from RICHS[244]. Briefly, eQTMs were calculated using linear models in MatrixEQTL[302], adjusted for fetal sex, 5 gene expression PCs and Planet-estimated cell types, considering all the genes in a 0.5 Mb window upstream and downstream of each CpG. Out of the 28 and 214 SMR-significant DNAm sites in MDD and SCZ, we found that 3 and 43 correlated with the expression levels of 4 and 26 significant eQTM-genes (FDR < 0.05), respectively, supporting the placental function of the CpG sites identified in the overlap between the two analyses. The eQTMs can be found in Appendix 12.

Interestingly, out of the 26 eQTM-genes related to SCZ, 10 mapped to the HLA region and among these, 6 were correlated with DNAm levels of more than one CpG, with 91 significant combinations between 6 and 21 unique genes and CpG sites, respectively. In particular, the expression levels of both *TOB2P1* and *ZKSCAN4* were correlated with the DNAm of 21 and 18 CpG sites pleiotropically associated to SCZ, respectively, confirming the high complexity of the locus. Outside the HLA region, two eQTM-genes correlated with the DNAm levels of more than one CpG site, namely *SFMBT1* and *VPS37B*. Specifically, the expression of *SFMBT1* on chromosome 3 was correlated with DNAm at 4 different contiguous loci that were independently identified in the SMR approach, while *VPS37B* on chromosome 12 correlated with the DNAm levels of two different CpG sites, with the same associated top SNP in SMR. In conclusion, our results suggest some regional pleiotropy, also outside the complex HLA region, that could have a functional impact in the placental expression landscape.

Next, we identified the intersect between SMR, colocalization and eQTM results for each trait (Figure 13). In the case of MDD, one placental DNAm site colocalized with an MDD-associated genomic locus on chromosome 14, and DNAm levels correlated with the placental expression of *LRFN5*. In SCZ, 40 placental DNAm sites colocalized with 12 SCZ-associated loci on chromosomes 1, 3, 6, 7, 8, 11, 12, 18, 19 and 22, with DNAm levels that correlated with the placental expression of 23 nearby genes (*GLB1L3, H2BC6, H3C4, KCNG2, LETM2, NAGA, PGBD1, PSMG3, RNF39, SDCCAG8,*

SFMBT1, SLC6A16, TOB2P1, TRIM27, TXNL4A, VARS2, VPS37B, ZKSCAN4, ZNF165, ZNRD1-AS1, ZSCAN12, ZSCAN12P1, and *ZSCAN23*).

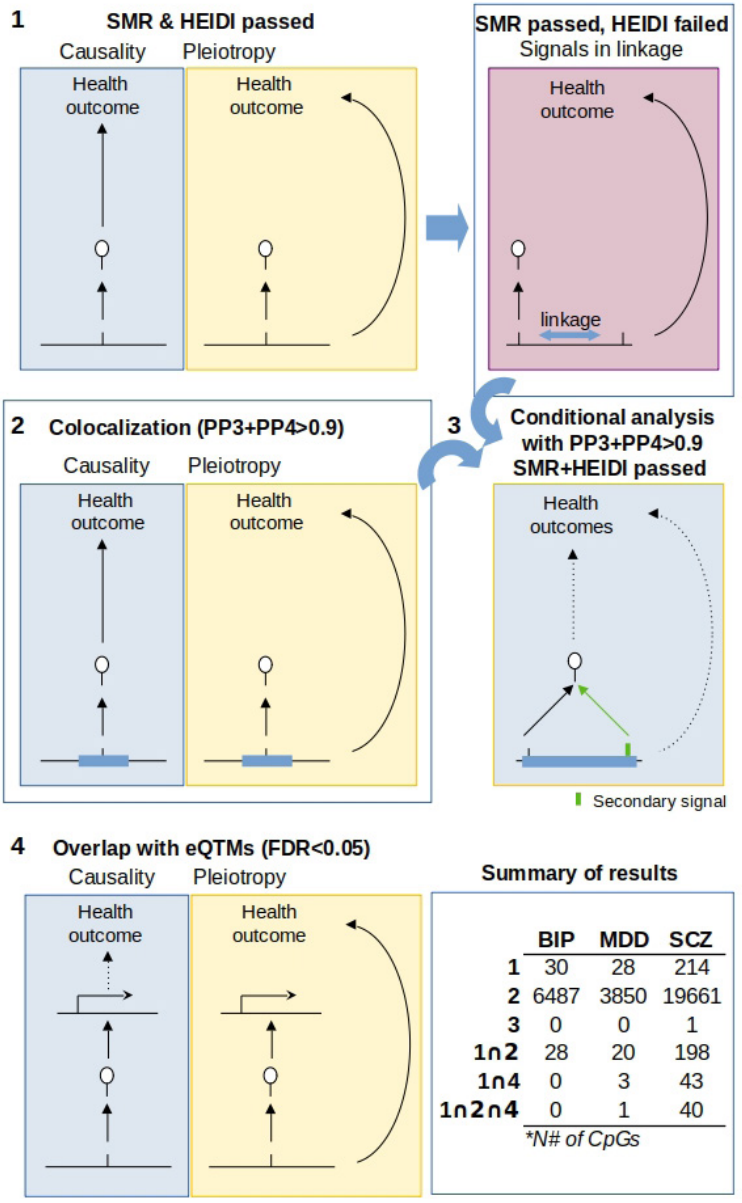


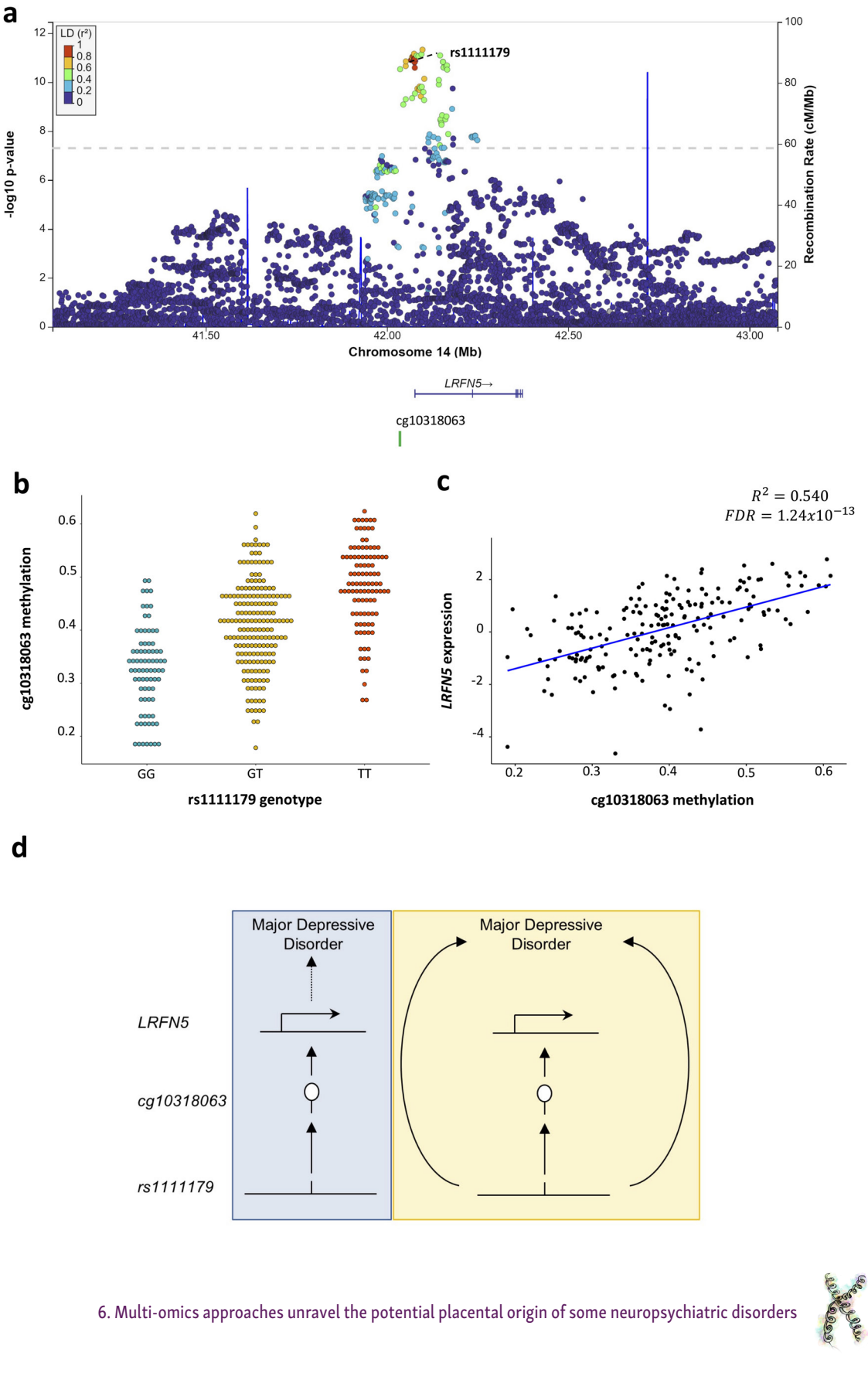
Figure 13. Workflow and results of the main analyses performed.

In a first step, we performed the SMR and HEIDI tests, followed by colocalization. Those hits that passed SMR but failed to pass HEIDI and were suggestive of association with both traits according to colocalization were considered for a conditional analysis with a second round of SMR and HEIDI tests, with the aim of discovering secondary associated signals. Finally, those CpG sites passing the SMR and HEIDI tests were interrogated in the eQTMs database of the RICHS cohort. The CpG sites overlapping the four approaches is summarized in the right-bottom table.

The overlapping hit in MDD was cg10318063 (Bonferroni P_{SMR} = 0.003) that colocalized with the MDD-associated locus on chr14:41940872-42476274 (PP3+PP4 = 1) (Figure 14). The DNAm levels of this CpG site correlated with the expression of *LRFN5* in placenta (FDR = $1.23x10^{-13}$ and R^2 = 0.540). Remarkably, several SNPs in *LRFN5* have been found to be pleiotropically associated with both MDD and chronic pain, although the effector tissue or cell type have not been fully established[303]. On the other hand, it is also known that this gene is expressed in TB stem cells, and therefore could play a role in placenta[304]. Finally, the CpG site identified is located in the promoter region of *LRFN5*. Altogether, these results suggest that the reported association is potentially causal, rather than pleiotropic.

Figure 14. A CpG site pleiotropically associated with MDD. ▶

The mVariant rs1111179, highlighted as a purple diamond, with the -log10 P values of the original MDD GWAS is represented in the **a**, locusZoom plot. The X-axis displays the involved genomic region of chromosome 14 in Mb, showing the distribution of the coding genes in the locus, as well as the location of the cg10318063 mSite. The Y-axis shows the -log10 P value from the original GWAS, and the SNPs are color coded as a function of the LD with the highlighted SNP considering the 1000G reference panel. The mQTL, rs1111179-cg10318063 ($P_{nominal}$ = $2.446x10^{-34}$) is plotted in the **b** dotplot, where the Y-axis represents the beta DNAm values of the indicated mSite, ranging from 0 to 1. The X-axis displays the genotype of the mVariant. The eQTM of cg10318063 is portrayed in the **c** dotplot, where the X-axis represents the DNAm values of the CpG site involved, ranging from 0 to 1. The Y-axis displays the expression values of the *LRFN5* gene in placenta. The hypothesis of the pleiotropical association between the SNP, the DNAm values of the CpG site in placenta, the gene expression levels of *LRFN5* in placenta and MDD are schematically represented in **d**. The vertical pleiotropy (or causal association) hypothesis is represented with a blue background, and the horizontal pleiotropy hypothesis is highlighted with a yellow background.

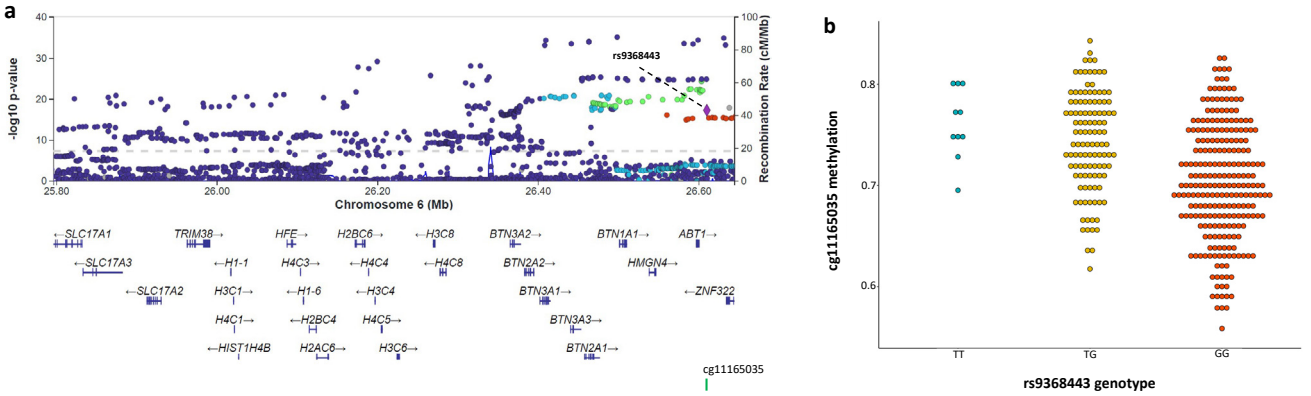

a
LD (r²)
1
0.8
0.6
0.4
0.2
0
rs1111179
-log10 p-value
Recombination Rate (cM/Mb)
Chromosome 14 (Mb)
41.50
42.00
42.50
43.00
LRFN5→
cg10318063
b
cg10318063 methylation
GG
GT
TT
rs1111179 genotype
c
$R^2 = 0.540$
$FDR = 1.24x10^{-13}$
LRFN5 expression
cg10318063 methylation
d
Major Depressive Disorder
Major Depressive Disorder
LRFN5
cg10318063
rs1111179

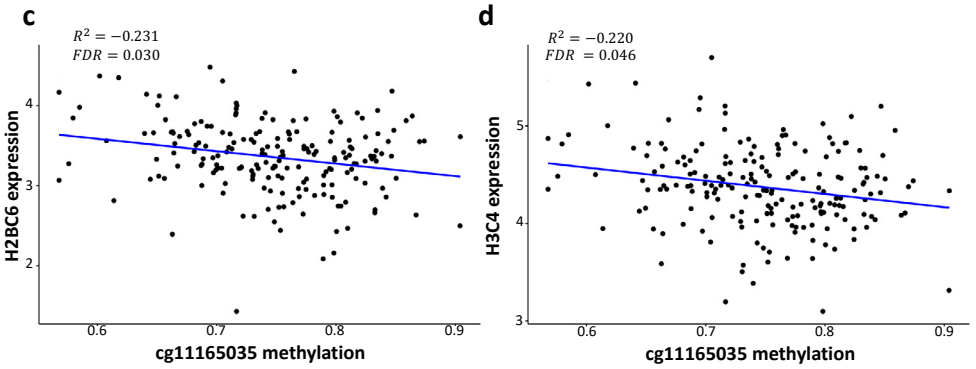
Among the CpG sites found in SCZ, cg11165035 was pleiotropically associated with the disorder (Bonferroni P_{SMR} = 0.043), colocalized with the GWAS peak on chr6:24933869-30563869 (PP3+PP4 = 0.999), and the DNAm levels of the CpG site were correlated with placental expression of *H2BC6* and *H3C4* (FDR = 0.030 and R^2 = -0.231, and FDR = 0.046 and R^2 = -0.220, respectively) (Figure 15). These two histone-coding genes are located on chromosome 6, outside the HLA region but close to it. Importantly, a relatively recent paper described a regulatory complex of different non-coding RNAs that inhibits the expression of *H2BC6* in the placenta of mothers with unexplained recurrent spontaneous abortions, provoking an enhanced apoptosis in TBs[305]. In addition, it has been shown that alcohol exposure during early neurulation can induce aberrant changes in DNAm patterns in mice embryos, with associated changes in gene expression, including *H3C4*[306].

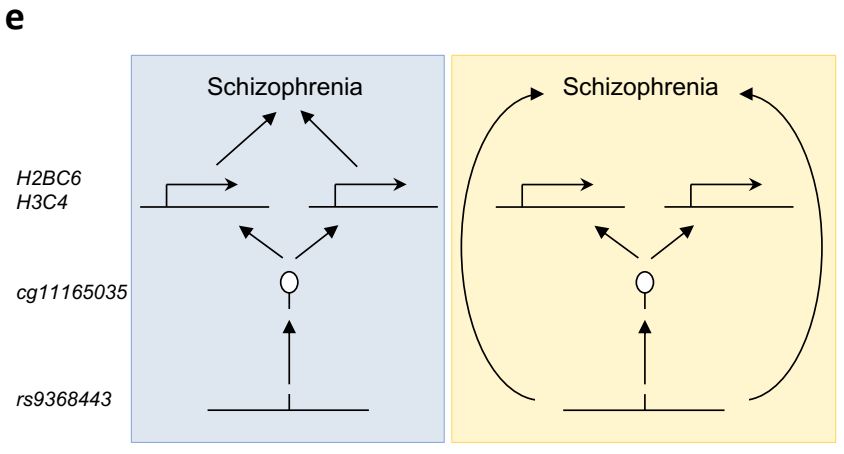
Finally, we performed a Reactome gene set analysis[296] of the 26 genes at the intersection between the SMR and eQTM results in SCZ. We observed enrichment for 101 gene sets, including many epigenetic regulatory pathways, such as different histone modifications, DNAm, and chromatin condensation and organization, suggesting that SCZ risk alleles could change the epigenetic landscape of placenta through the regulation of, among others, the aforementioned histone-coding genes. Additionally, we observed a remarkable enrichment in immune-related pathways such as human cytomegalovirus (HCMV) infection and overall viral infection, reinforcing the idea of MIA as a link between placenta and SCZ risk. The complete Reactome results can be found in Appendix 13.

Figure 15. One CpG site pleiotropically associated with SCZ is associated to placental expression of two different histone-coding genes. ▶

The mVariant rs9368443, highlighted as a purple diamond, is shown in the original SCZ GWAS in the **a** locusZoom plot. The X-axis displays the genomic region involved in chromosome 6 in Mb, showing the distribution of the coding genes in the locus, as well as the location of the mSite cg11165035. The Y-axis shows the -log10 P value from the original GWAS, and the SNPs are color coded as a function of the LD with the highlighted SNP considering the 1000G reference panel. The mQTL, rs9368443-cg11165035 ($P_{nominal}$ = $2.054x10^{-11}$) is plotted in the **b** dotplot, where the Y-axis represents the DNAm beta values of cg11165035, ranging from 0 to 1. The X-axis displays the genotype of the mVariant involved. The eQTMs of the significant mSite are portrayed in the **c** and **d** dotplots, where the X-axes show the DNAm values of the cg11165035 CpG, ranging from 0 to 1. The Y-axes display the expression values of *H2BC6* and *H3C4* genes in placenta. The hypothesis of the pleiotropical association between the SNP, the DNAm values of the CpG sites in placenta, the gene expression levels of *H2BC6* and *H3C4* in placenta and SCZ are schematically represented in **e**. The vertical pleiotropy (or causal association) hypothesis is represented with a blue background, and the horizontal pleiotropy hypothesis is highlighted with a yellow background.


a
-log10 p-value
Recombination Rate (cM/Mb)
rs9368443
Chromosome 6 (Mb)
25.80
26.00
26.20
26.40
26.60
←SLC17A1
TRIM38→
HFE→
H2BC6→
←H3C8
BTN3A2→
BTN1A1→
ABT1→
←SLC17A3
←H1-1
H4C3→
←H4C4
←H4C8
BTN2A2→
HMGN4→
←SLC17A2
H3C1→
←H1-6
←H3C4
BTN3A1→
←ZNF322
H4C1→
←H2BC4
H4C5→
BTN3A3→
←HIST1H4B
H2AC6→
H3C6→
BTN2A1→
cg11165035
b
cg11165035 methylation
TT
TG
GG
rs9368443 genotype


c
$R^2 = -0.231$
$FDR = 0.030$
H2BC6 expression
cg11165035 methylation
d
$R^2 = -0.220$
$FDR = 0.046$
H3C4 expression
cg11165035 methylation

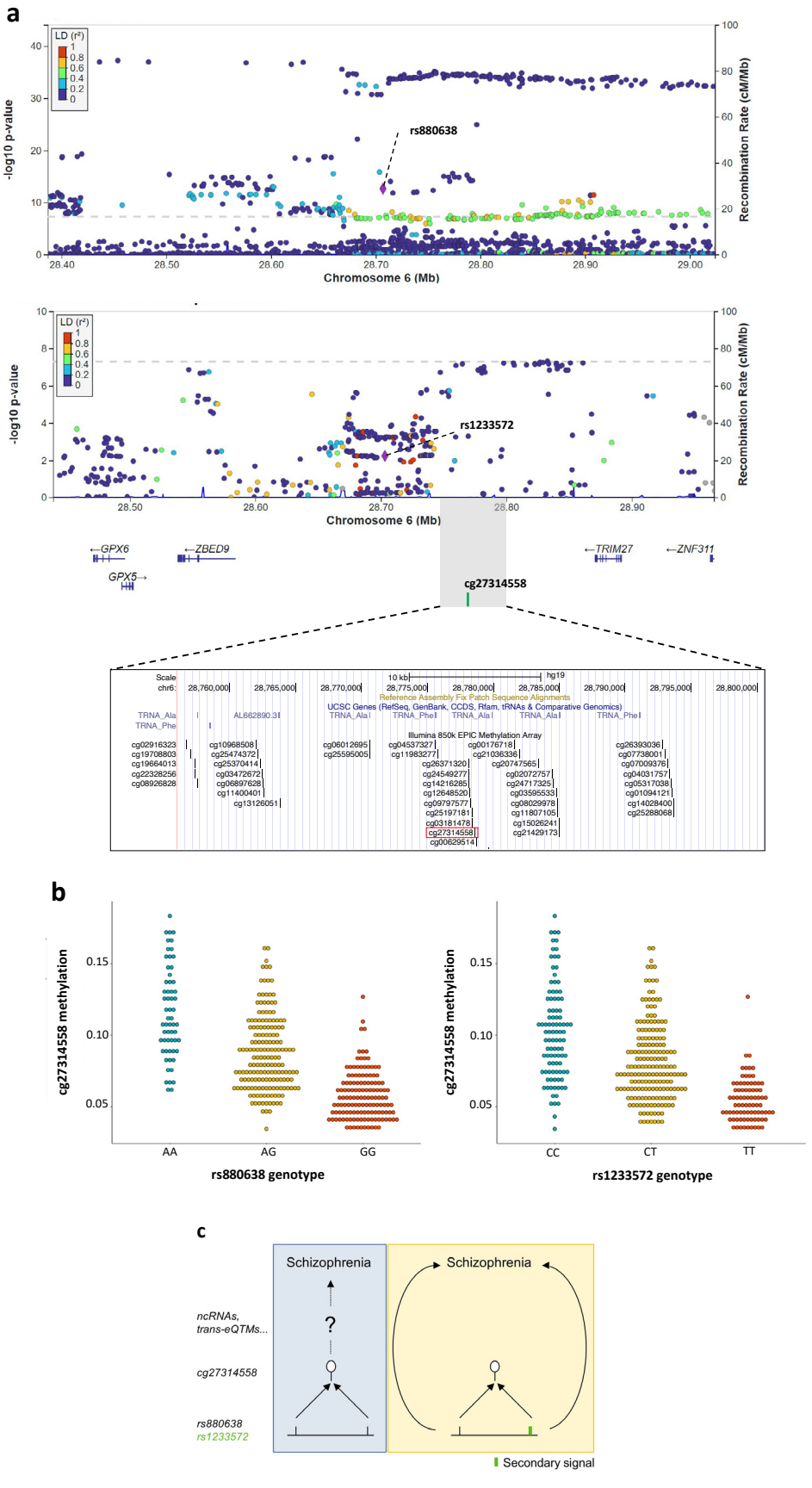

e
Schizophrenia
Schizophrenia
H2BC6
H3C4
cg11165035
rs9368443

7. A secondary potentially causal signal is detected in schizophrenia

The developers of the SMR software designed the HEIDI test assuming that a single causal variant in the *cis*-mQTL region affects both DNAm and the trait analyzed[121]. Under the assumption of pleiotropy, when there are multiple causal variants in a region, the pleiotropic signal of one causal variant will be diluted by that of other non-pleiotropic causal variants. We therefore performed a GCTA conditional analysis[293] conditioning for the top associated *cis*-mQTL in both the GWAS and mQTL data sets. In those cases where a secondary signal was pinpointed by the presence of heterogeneity (P_{HEIDI} < 0.05) in the *cis*-mQTL region, in either the GWAS or the mQTL data set, we performed another round of conditional analyses conditioning only on the secondary signal in both the GWAS and mQTL data sets, and then reran the SMR and HEIDI test at the top *cis*-mQTL using the estimates of SNP effects from the conditional analyses. We applied this approach to those CpG sites that passed the SMR test but failed to pass the HEIDI test. Nevertheless, due to power constrains, only those CpG sites with suggestive support for association with both traits according to Coloc (PP3+PP4 > 0.9) were considered.

Figure 16. One CpG is pleiotropically associated with SCZ and marked by two independent signals. ▶

The two mVariants rs880638 and rs1233572 highlighted as purple diamonds, are shown in the original (top) and the conditional (bottom) SCZ GWAS in the **a** locusZoom plot. The X-axis displays the genomic region involved in chromosome 6 in Mb, showing the distribution of the coding genes in the locus, as well as the location of the mSite cg27314558. The Y-axis shows the -log10 P value from the original and conditional GWAS, and the SNPs are color coded as a function of the LD with the highlighted SNP considering the 1000G reference panel. Additionally, the UCSC region with UCSC genes and Illumina EPIC Methylation Array probes from region chr6:28,760,00-28,800,000 can be found in the same panel. cg27314558 has been highlighted in red. The two mQTLs, rs880638-cg27314558 ($P_{nominal}$ = $4{,}57x10^{-50}$) and rs1233572-cg27314558 ($P_{nominal}$ = $3{,}263x10^{-30}$), are plotted in the **b** dotplots, where Y-axes represent the cg27314558 DNAm beta values, ranging from 0 to 1. The X-axis displays the genotype of the corresponding mVariant. The hypothesis of the pleiotropical association between the SNPs, the DNAm values of the CpG sites in placenta, and SCZ are schematically represented in **d**. The vertical pleiotropy (or causal association) hypothesis is represented with a blue background, and the horizontal pleiotropy hypothesis is highlighted with a yellow background.


a
LD (r²)
1
0.8
0.6
0.4
0.2
0
-log10 p-value
Recombination Rate (cM/Mb)
rs880638
Chromosome 6 (Mb)
rs1233572
←GPX6
←ZBED9
GPX5→
←TRIM27
←ZNF311
cg27314558
Illumina 850k EPIC Methylation Array
b
cg27314558 methylation
AA
AG
GG
rs880638 genotype
CC
CT
TT
rs1233572 genotype
c
Schizophrenia
ncRNAs,
trans-eQTMs...
cg27314558
rs880638
rs1233572
Secondary signal

We discovered a secondary hit in SCZ, in which two independently associated SNPs, i.e. rs880638 (the primary signal) and rs1233572 (the secondary signal) had a pleiotropic effect on both cg27314558 and SCZ (Bonferroni P_{SMR} = 0.015 and P_{SMR} = 0.024, respectively) (Figure 16). Even though we did not find any significant eQTM for this CpG, located in the HLA region, we must bear in mind that small RNAs were not included in the eQTM analysis. Having a closer look to this genomic region using the UCSC genomic browser[307], we found that cg27314558 is surrounded by multiple transfer RNAs (tRNA), an alanine tRNA (*TRA-AGC6-1*) being the closest (Figure 16). In this context, it is well known that impaired placental amino acid transfer is associated with fetal growth restriction[308]. Additionally, a recent study revealed that placentas of preterm births show a significant negative correlation with alanine concentration[309]. Our results, namely the presence of two independent signals pointing to a single DNAm site in placenta that moreover, is close to multiple aminoacyl-tRNAs, suggest a causal association of the CpG site with SCZ. The conditional GWAS and mQTLs, as well as the results of the conditional SMR analysis can be found in Appendix 14.

8. Detection of pleiotropic associations between bipolar disorder and schizophrenia, and STB- and TB-interacting mQTLs

We filtered the imQTLs by $P_{nominal} < 5x10^{-8}$, as suggested by the developers of SMR. We then classified the cell type-imQTLs as positive or negative according to the article by Kim-Hellmuth *et al.*, with the aim of retrieving potentially cell type-specific mQTLs (that is, positive imQTLs) rather than those that more likely reflect interactions with another cell population that is decreasing (that is, negative imQTLs). In summary, positive imQTLs show an increase of the main genotype effect from low to high cell proportions, i.e., the mQTL effect size increases with the cell type abundance. Meanwhile, negative imQTLs show a decrease of the mQTL effect size as a function of the cell type proportion.

We used SMR to combine positive imQTLs with the BIP, MDD and SCZ GWAS, and obtained one STB-imQTL (related to BIP), and two TB-imQTLs (one each in BIP and MDD) (Appendix 15). We found an imQTL involving rs11637580 and cg27130493 (Figure 17) in which cg27130493 appeared to be associated with BIP and changed its DNAm levels as a function of TB proportion, in a rs11637580 genotype-dependent manner. This CpG site is located in the gene body of *SMAD3*. The overexpression of this gene in placenta has been described to activate the ability of TBs to form endothelial-like networks, while its defect has been associated with preeclampsia[310]. Additionally, it is a well-known target for lithium treatment in BIP[311]. Interestingly, rs11637581 was not directly associated with BIP in the original GWAS, suggesting a possible causal association between placental DNAm and BIP at this CpG site.

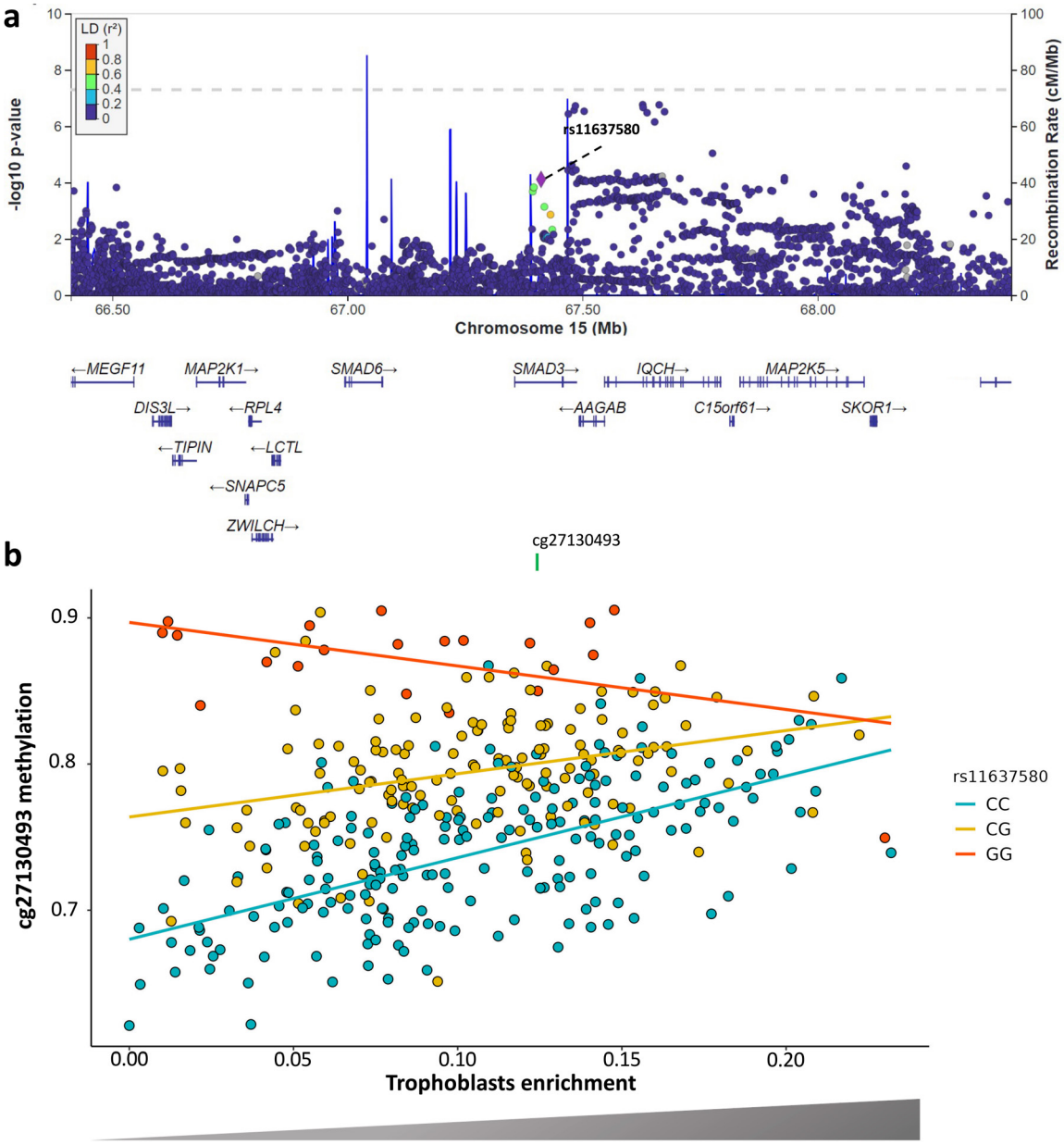


Figure 17. TB-imQTL pleiotropically associated with BIP.

The imVariant rs11637580, highlighted as a purple diamond, is shown in the original BIP GWAS in the **a** locusZoom plot. The X-axis displays the involved genomic region in chromosome 15 in Mb, showing the distribution of the coding genes in the locus, as well as the location of the imSite cg27130493. The Y-axis shows the -log10 P value from the original GWAS, and the SNPs are color coded as a function of LD with the highlighted SNP considering the 1000G reference panel. The TB-imQTL ($P_{nominal}$ = 1.671x10^{-8}) is pictured in the **b** dotplot. The X-axis represents the cg27130493 DNAm beta values and the Y-axis the TB proportion, both ranging from 0 to 1. The genotype of the rs11637580 imVariant is color coded as indicated in the legend.

9. Most of the pleiotropic associations are placenta-specific and are not covered by brain and fetal brain studies

In order to ascertain the tissue specificity of our findings, we also performed the same SMR analyses with fetal brain and brain *cis*-mQTLs in BIP, MDD and SCZ. The fetal brain *cis*-mQTL database was published by Hannon *et al.* in 2015[84]. Briefly, mQTLs were calculated in 166 human fetal brain samples (56-166 days post-conception). The DNAm data were obtained with the Illumina Infinium HumanMethylation 450K array, and mQTL mapping was performed with the MatrixEQTL R package[302], considering a *cis*-window of ±0.05 Mb. Only mQTLs with a $P < 1.5x10^{-9}$ were made available, resulting in 556,513 fetal brain *cis*-mQTLs that were downloaded from the SMR portal. The second brain *cis*-mQTL database was originally from a 2018 publication by Qi *et al.*[312], in which *cis*-mQTLs from three different data sets were meta-analyzed: 468 brain cortical region samples from the Religious Order Study and the Memory and Aging Project[88], 166 fetal brain samples from the study by Hannon *et al.*, and 526 frontal cortex region samples[313] were included, amounting to a final sample size of 1,160 brain samples and nearly 6M *cis*-mQTLs. As in the case of the fetal brain mQTL database, this data set was downloaded from the SMR data portal.

In BIP we obtained 13 and 35 significant SMR hits (Bonferroni $P_{SMR} < 0.05$ and $P_{HEIDI} > 0.05$) in the fetal brain and brain mQTL data sets, respectively, and with same criteria, we obtained 10 and 23 SMR hits in MDD, and 50 and 188 in SCZ. The results can be found in Appendix 16. As shown in Figure 18, the trait with the largest overlap of pleiotropically associated DNAm sites among tissues is SCZ, although the overlap is very limited in all the traits, also at the level of mVariants. In the case of the fetal brain database, the small sample size limits the number of mQTLs detected and could lead to underestimate the proportion of the genetic risk that acts through the DNAm of the fetal brain. When we compared our most reliable hits in placenta (intersection

among SMR, Coloc and eQTM results) with the brain SMR results, neither MDD nor SCZ presented any overlapping DNAm sites.

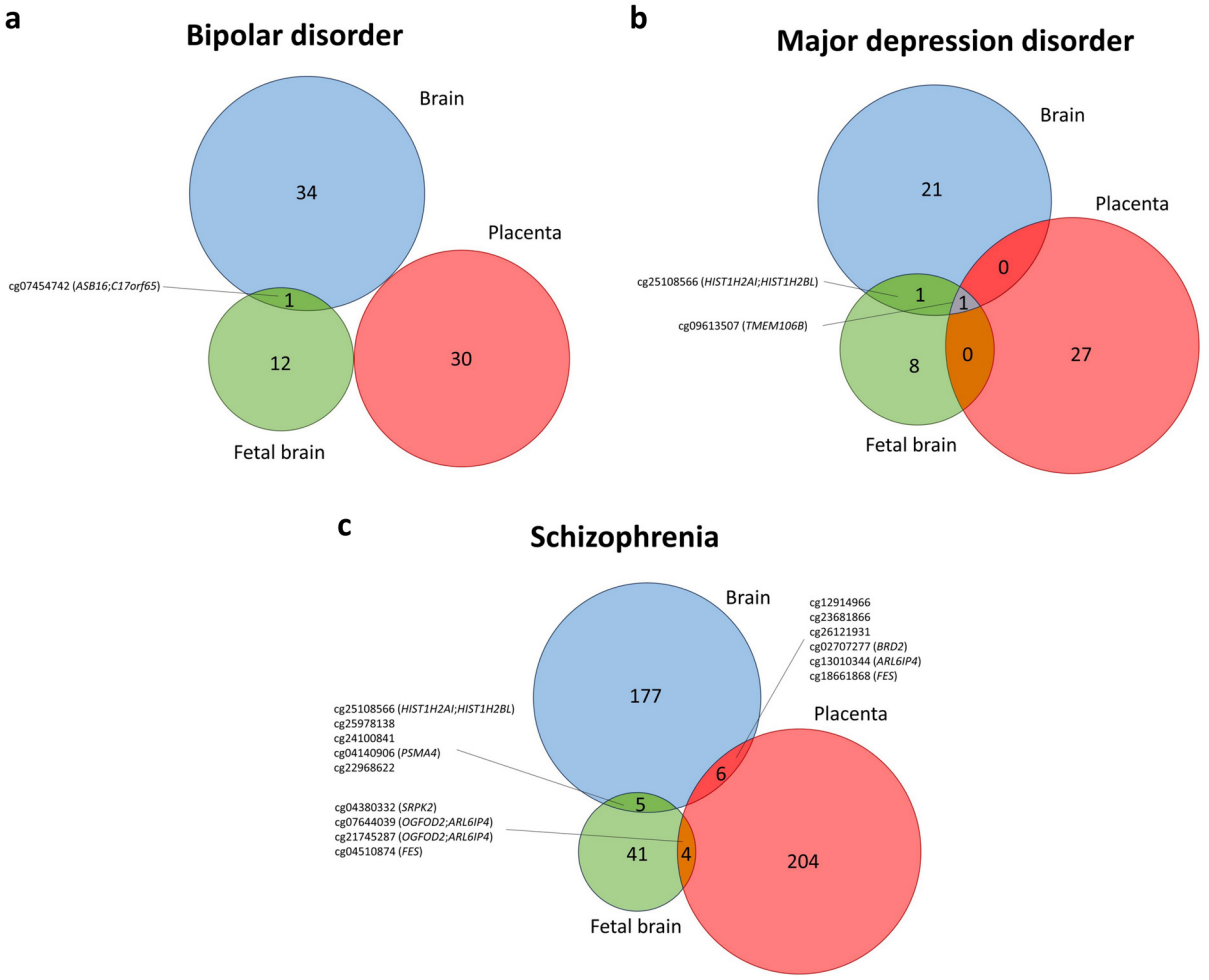


Figure 18. Overlap between the SMR results in brain, placenta and fetal brain, in BIP, MDD and SCZ.

The overlap between the mSites pleiotropically associated between the three tissues and BIP, MDD and SCZ are represented in the **a**, **b** and **c** Venn diagrams, respectively. Overlapping mSites are shown, and also the closest gene from the Illumina annotation file (between brackets).

DISCUSSION

1. Maternal ppBMI EWAS

As far as we know, this is the largest EWAS meta-analysis conducted to date on placental DNAm. We have analyzed a total of 2,631 mother-child pairs from 10 different PACE cohorts from Europe, North America, and Australia. We have identified 27 CpG sites associated to maternal ppBMI, some of which showed up to 2% lower DNAm per 10-unit higher BMI. Although such a difference in an individual woman is unlikely in the context of a pre-pregnancy intervention, we consider that it could represent the difference between women in the normal range of BMI and women in the obesity category.

The most significant association was observed for cg08219219, located in the eighth exon of *EPHX3*, for which a ppBMI difference of 10 units is associated with a 1.1% lower placental DNAm. It has been shown that soluble epoxide hydrolases such as *EPHX3* have higher activity in obese mice[314]. Additionally, it has been suggested that this family of hydrolases could act as a therapeutic target for metabolic and cardiovascular abnormalities related to obesity[315]. In turn, we highlighted two other significant hits showing the largest positive and negative effect coefficients. The cg14704941, in the first intron of *CSRP3*, presented with 2% higher placental DNAm per 10-unit ppBMI. *CRSP3* knockout mice develop dilated cardiomyopathy with hypertrophy and heart failure after birth[316]. The effect-coefficient of cg04724807, located upstream of *SYT16*, represented about 1.8% lower placental DNAm per 10-unit greater ppBMI. *SYT6* is over-expressed in pancreatic islet cells in response to high glucose challenge and is thought to play a role in insulin secretion[317]. As previously stated, maternal obesity has been described to be associated with obesity, diabetes and cardiometabolic conditions in offspring later in life[69,70]. The fact that our EWAS identified CpG sites near those metabolically relevant genes highlights the plausibility that they may play a role in the link between maternal obesity and future health outcomes in children.

Among our significant signals, we also found two CpG sites, cg00423969 and cg14163484, 1.5 kb upstream of the *FER1L5* promoter, presenting lower placental

DNAm levels associated with higher maternal ppBMI. Remarkably, *FER1L5* encodes a dysferlin- and myoferlin-related protein, which has been predicted to have a role in vesicle trafficking and muscle membrane fusion events[318]. Both vesicle trafficking and membrane fusion are crucial events in placental development, since they allow the formation of the STB, the uninterrupted and multi-nucleated mass that covers the villous trees and enables the interplay with the mother[319]. In addition, *CMIP* and *GPX1*, two of the genes annotated to maternal ppBMI-associated CpG sites, may present relevant biological roles in pregnancy. For example, different CpG sites surrounding *CMIP* have been associated with preeclampsia in a placental DNAm study[320]. *GPX1* is an antioxidant gene, and its mRNA levels are lower in the placenta of obese mothers compared to normal-weight mothers[321]. Finally, in the context of obesity, *LGR4*, another gene identified in the present study, bears an activating variant that contributes to abdominal visceral fat accumulation and therefore, to central obesity[322], suggesting that both genetic and epigenetic regulation at this locus may play a role in obesity-related phenotypes.

Regarding enrichment analyses, one of the most interesting findings is that several significant CpG sites are located close to cancer-related genes. It has been recently described that the placenta is organized as a big mass of tumoral clones, with rapidly occurring cell divisions that enable selection of good cells that will eventually form the baby. Additionally, both cancer and placental STB are invasive tissues with many biological parallelisms[323,324]. Indeed, it is not surprising that factors that are relevant to the placenta, such as maternal obesity, could affect genes that are relevant to cancer. The other pathway that was enriched for altered genes is oxidative stress. It is well known that excessive accumulation of fat mass is linked to oxidative stress. Moreover, peroxisomal fatty acid oxidation seems to be enhanced in the placenta of obese women, while mitochondrial activity is impaired, with a greater lipid storage and an altered transfer of lipids to the fetus[325]. Altogether, there is growing evidence suggesting that obesity-induced oxidative stress is a crucial factor involved in the risk for adverse outcomes in pregnancy[326,327].

Another interesting finding that deserves further investigation is the observation that differentially methylated CpG sites are enriched for *ZNF217* binding sites. This TF

is epigenetically altered in placental cells under hypoxia[328], and it has been suggested that maternal obesity during pregnancy causes placental hypoxia[329]. Additionally, Delahaye *et al.*, in one of the previous placental mQTL studies, already reported an enrichment of *ZNF217* on their eQTMs, and suggested that the alteration of the binding of this TF is predicted to disturb the formation of a gene repression complex and thus, lead to an increase of expression in placenta[105]. However, whether this TF can drive the methylation machinery to selected regions of the genome and cause epigenetic changes has not been demonstrated yet. Similarly, the overlap between our CpG sites and several BW-associated regions suggests that both fetal genetic and placental epigenetic factors may contribute to the regulation of fetal growth. But none of our overlapping CpG sites were reported in the study by Tekola-Ayele *et al.*, a placenta multi-omics integration study that identified several candidate functional genes for BW. Thus, further research will be required to completely clarify this issue.

Our study has notable strengths but also several limitations. As previously mentioned, we have been able to coordinate a large number of cohorts and thus obtain an important sample size. Additionally, we have the expertise from previous works in which robust pipelines were implemented for EWAS, and additionally, we have run both the quality control protocol and meta-analysis in two independent institutions. Finally, none of the maternal ppBMI-associated CpG sites form the current study are among the problematic probes with absolute methylation differences greater than 10% between Illumina 450K and EPIC arrays that had been identified in a previous study from our own group[330].

Regarding limitations, we did not have access to individual data addressing whether each ppBMI measurement was self-reported or taken at the end of the first trimester of pregnancy. Therefore, we cannot use this variable as a covariate nor compare between measurement types. We are very aware that self-reported ppBMI may not be the most accurate measurement for our variable of interest, as may also be the case of the measurement of BMI at the end of the first trimester. However, self-reported ppBMI has been shown to be reliable and highly correlated to measured BMI at 12 weeks of gestation ($r = 0.96$, $P < 0.0001$)[331]. Second, the unavailability of genotype data in some of the participating cohorts did not allow us to add genotype

PCs to our models, and there might be residual confounding by population structure that we did not account for. Third, most of the cohorts were composed of a majority of individuals of European descent, which limits generalizability of our findings to other populations.

Additionally, we are aware that RefFreeCellMix, the R package employed for adjustment of cell mixtures, is a PCA-type correction method, and therefore presents the risk of over-correcting the results, especially in dense signal scenarios like the Illumina BeadChips, due to the capture of the signal by some of the top components of the estimation[332]. This, together with the fact that the Bonferroni correction is very strict and that our approach does not consider the correlation between nearby CpG sites, may have left some discoveries out of focus. However, we have preferred to be strict and report only the most robust results.

In summary, here we present the largest EWAS of placental DNAm performed to date. We identify 27 CpG sites at which we observe placental DNAm variations of 0.5–2.0% by 10-unit maternal ppBMI difference. Additionally, our DNAm findings seem to be placenta-specific, showing minimal overlap with a previous meta-analysis in cord blood DNAm in relation to maternal ppBMI. The differentially methylated CpG sites are mainly located in open sea regions, with a complete depletion from CpG islands, and enriched in cancer- and oxidative stress-related pathways. These observations, together with the fact that maternal ppBMI is associated with placental DNAm at CpG sites located close to obesity-related genes, leads us to hypothesize that placental DNAm could be one of the mechanisms by which maternal obesity is associated with aberrant fetal growth and maybe, other metabolic health outcomes in offspring later in life. However, we cannot rule out that the changes observed could be markers of exposure to high ppBMI and therefore, our findings will need to be supplemented by functional studies or causal inference analyses to better understand if they truly have a role in pregnancy complications or long-term metabolic outcomes.

2. Placental *cis*-mQTLs

Regarding the placental mQTL analysis, we have constructed different placental *cis*-mQTL databases with 368 placenta samples from the INMA project. Importantly, we have made all results publicly available both in their raw formats as well as on a user-friendly, shiny-based browser that allows to search for the mQTLs of interest by CpG, SNP and/or genomic coordinates. We believe that this tool will be useful to the scientific community, and it is available at the following web address: https://irlab.shinyapps.io/shiny_mqtl_placenta/

In general, the placental *cis*-mQTLs were depleted in regions that are usually hypomethylated, such as promoters and CpG islands, highly likely due to the fact that lower and more stable DNAm values will hardly correlate with the genotype or any other variable. In turn, an enrichment of intermediate DNAm values was observed, usually present in CpG island shelves and shores, as well as in open sea regions. The genotype-regulated placental DNAm seems highly placenta-specific and interestingly, it is enriched in placenta-specific active chromatin marks, and in pathways such as neuropathy, mood disorders and BIP, among others. This, together with the fact that mQTLs have recently been highlighted as powerful instruments to reveal molecular links to traits otherwise missed by eQTL-GWAS colocalization approaches[90], encouraged us to conduct a multi-omics study to try to unveil the placental contribution to the developmental origins of different neuropsychiatric disorders.

Kim-Hellmuth *et al.* showed that cell type-ieQTLs are enriched in GWAS signals and can improve GWAS-eQTL matching for the mechanistic understanding of those loci[91]. In this context, we wondered whether this could also be true for placental mQTLs and calculated STB- and TB-imQTLs. STBs were considered for several reasons, 1) they are the most abundant cell type in term placenta[43], 2) they cover the entire surface of the villous trees in the placenta, being in direct

contact with maternal blood[44], 3) they orchestrate the complex biomolecular interactions between the fetus and the mother, and 4) they act as an important endocrine organ, producing numerous growth factors and hormones that support and regulate placental and fetal development and growth[45-48]. Given that TBs are the progenitor cells of STBs, and that there is a very strong negative correlation between the proportions of the two cell types in our samples, we decided to calculate TB-imQTLs as well. We observed higher statistical power for the most abundant cell type and therefore, a larger amount of STB-imQTLs compared to TB-imQTLs. However, there was a remarkable sharing and, as expected, the allelic effects in the overlapping imQTLs were negatively correlated between the two cell types. As TBs are known to differentiate into STBs throughout gestation[45], and STB content in term placenta is positively correlated with GA at birth, we wanted to know whether cell type-imQTLs are equivalent to GA-imQTLs. This was not the case, revealing that TB-to-STB differentiation has a notable effect on placental DNAm that is not explained by GA.

Remarkably, placental DNAm is pleiotropically associated with BIP, MDD, and in particular with SCZ, while it seems to barely associate with early onset conditions such as ADHD and ASD. These results could arise from the sample sizes and power constraints of the GWAS available, as well as from the higher heritability of certain disorders, including BIP and SCZ. However, it is important to note that a recent article found that MDD shows a very high polygenicity compared to other psychiatric disorders; that is, more genetic variants with weaker effects contribute to the overall genetic signal in MDD and make the trait less annotable[333]. In contrast, ADHD, BIP and SCZ showed the highest discoverability and hence a more annotable genetic signal. Subsequently, the estimated sample size required to reach 90% SNP heritability was more than eight times larger for MDD than for ADHD, BIP and SCZ. Therefore, the moderate signal in ADHD and even more remarkably, the presence of a considerable pleiotropy between placental DNAm and MDD, suggest that our findings are guided not only by the strength and annotability of the genetic basis of the diseases studied, but rather by a genuine association with placental DNAm. In addition, it is well known

that BIP, MDD and SCZ share a common genetic background and therefore, it is plausible that part of the genetic risk could act through common processes at the same developmental stages[334]. In summary, these observations, together with the correlations observed between the number of SMR hits and the GWAS sample sizes and significant loci, suggest that although the genetic basis seems to be crucial, there is a genuine contribution from the placental DNAm to those neuropsychiatric disorders.

It is possible to gain insight into the pathogenesis of complex disorders by defining the environmental, biological and temporal context in which genes increase disease susceptibility. In 2018, Ursini *et al*. discovered that when ELC are present, the polygenic risk score (PRS)-explained risk of SCZ is more than five times higher than when they are absent[210]. SCZ loci that interact with ELC are not only highly expressed in placenta, but also differentially expressed between complicated and normal pregnancies and enriched in response-to-stress pathways. Remarkably, three out of the genes whose expression correlates with DNAm sites associated with SCZ according to the different approaches presented here, were among the 248 placenta-enriched SCZ genes described by Ursini *et al.* (i.e. *NAGA*, *PSMG3* and *SFMBT1*). Therefore, we propose that those genes could mediate SCZ risk not only through placental expression, but rather through DNAm-regulated placental expression patterns.

More recently, placenta-specific SCZ-PRS has also been shown to interact with both brain volume and cognitive function, suggesting particular neurodevelopmental trajectories in the path towards SCZ[211]. In that work, Ursini and colleagues also studied the enrichment in placenta of genes involved in the risk to other neuropsychiatric disorders, and the relationship between neurodevelopmental outcomes and placenta-specific PRS for the same disorders. BIP presented a very high enrichment in genes that are expressed in placenta, although its placenta-specific PRS did not interact with neither brain volume nor cognition. The placental enrichment observed in BIP makes sense since SCZ and BIP present the highest genetic correlation compared to any other pair of psychiatric traits[335]. This is in line

with the convergence between the CpG sites associated with both BIP and SCZ found in this study. Nevertheless, there are remarkable differences in outcome between the two disorders, and SCZ patients have been described to be more prone to suffer from severe cognitive impairment than BIP patients[336]. In conclusion, placental DNAm could be important in BIP although through trajectories that are different from those characterized by Ursini and colleagues in SCZ. In any case, we believe that our findings support a novel hypothesis according to which placental DNAm could translate both the genetic basis and the environmental milieu into fetal genetic programs that could eventually result in impaired neurodevelopmental trajectories leading to SCZ, and maybe also to other neuropsychiatric disorders. This hypothesis will need to be tested with additional research.

Regarding the genes affected by placental DNAm that are pleiotropically associated with SCZ, we want to highlight that they were enriched mostly in epigenetic regulation pathways and immune-related terms, including HCMV infection and late events, as well as assembly of the human immunodeficiency virus (HIV) virion, among others. On the one hand, the multiple epigenetic pathways involved suggest that SCZ risk alleles could cause deep epigenetic changes in placenta as a result of DNAm modifications near the histone-coding genes *H2BC6* and *H3C4*. These genes have been previously related to recurrent spontaneous abortion and neurulation, respectively[305,306]. On the other hand, those histones, together with *VPS37B*, were some of the most relevant genes contributing to the enrichment in immune-related biological routes. This reinforces the idea of MIA being implicated in the neurodevelopmental origins of SCZ. Particularly for HCMV, it has been reported that the DNAm state of the host genome can make cells more or less prone to infection, and that histones participate in this process because, although HCMV is devoid of those proteins, it captures them from the host for the chromatinization of its own genome[337]. Moreover, it is known that congenital HCMV infection can cause long-term clinical manifestations such as hearing loss, neurodevelopmental disorders, ophthalmic complications, and ASD, among others[338]. The virus moves from mother to fetus by infecting TB in the placenta, altering the development and integrity of the organ, and eventually, inhibiting

placental cell differentiation, self-renewal, and migration. Last but not least, in a very recent transcriptome-wide association study by Ursini and colleagues, the placental genes and transcripts associated with SCZ risk were shown to be enriched for pathways related with the activation of the pathogenesis of both Coronavirus and influenza, further reinforcing the idea of the implication of viral infections[339].

Altogether, environmental insults that occur in pregnancy and early life, including maternal infections, are hypothesized to program the immune and developmental epigenetic code in the fetus, thereby influencing the risk to suffer from neurodevelopmental disorders later in life[340-342]. The MIA hypothesis proposes that exposure to a dysregulated maternal immune milieu *in utero* affects fetal neurodevelopment[130,343]. Moreover, in humans, maternal factors implicated in MIA are associated with epigenetic modifications in the placenta[344,345]. The placenta plays a pivotal role in maintaining immune homeostasis in the maternal-fetal interface. However, when a sustained placental inflammatory response occurs due to maternal environmental factors, the offspring can suffer from developmental abnormalities[346].

Apart from the genes leading the aforementioned pathway enrichments, several others are also worthy of mention, as noted by their correlation with multiple CpG sites associated with SCZ. On the one hand, among the genes in the HLA region and despite the high complexity of the locus, *TOB2P1* is the one locally correlated with more SCZ-associated DNAm sites, and one of those in the region correlated with a CpG site that colocalizes with SCZ according to Coloc. *TOB2P1* has been consistently prioritized in this study and additionally, its RNA expression level in placenta is regulated by the genotype of nearby SNPs[347]. On the other hand, outside the HLA region, we would like to mention *SFMBT1* and *VPS37B*. The former is an epigenetic regulator that is crucial for TB maintenance and placental development[348,349]. Moreover, it was one of the 139 placenta and SCZ-specific risk genes prioritized in the transcriptome-wide association study by Ursini and colleagues[339], and it is near a CpG site identified to be associated with the maternal

ppBMI in the previous EWAS. The relevance of *VPS37B* is beyond any doubt given its immune function and its involvement in viral infection. In conclusion, all those genes deserve further investigation.

Briefly, ASD and ADHD hits were cg24412801 and cg13666471, that lie within *KIZ* and *ST3GAL3* genes, respectively. On the one hand, *KIZ* encodes a protein localized to centrosomes, strengthening, and stabilizing the pericentriolar region, and has been found to boast several SNPs that colocalize with both ASD and DNAm from neonatal dried blood spots[350]. Up to date, this gene has not been described related to any pregnancy complication or placental biological pathways. On the other hand, *ST3GAL3* encodes a Golgi membrane protein that catalyzes the transfer of sialic acid and has been associated with intellectual disability. More importantly, it has already been reported as a risk gene for ADHD[280] that additionally is hypomethylated in the blood of individuals with the disorder[351]. In this study, it was shown that gene-environment interaction may moderate *ST3GAL3* expression, impacting epigenetic programming of brain development and maturation[351]. Up to date, this gene has not been described in the context of placental DNAm and the prenatal origins of ADHD. Further studies are required since none of these DNAm sites seem to regulate the expression of nearby genes in placenta and therefore, it cannot be ruled out the possibility of other tissues and organs being implicated in the observed associations.

Taken individually, the genes that are most likely to be causally involved in the different neuropsychiatric disorders studied were *LRFN5* in MDD, and *H2BC6* and *H3C4* in SCZ. In the case of *LRFN5*, several variants in this gene have been found to be pleiotropically associated with both MDD and chronic pain[303]. Additionally, it is well known that *LRFN5* is involved in brain cell communication and is located in a large and complex genomic niche that is highly conserved in mammals[352]. Specific locus structure of this region increases ASD susceptibility in males. In the case of the histone-coding genes on chromosome 6, a single SNP seems to pleiotropically associate with both SCZ and the DNAm levels of a CpG site that additionally correlates with the expression of those two genes in placenta. Together with the

fact that both genes play important roles in chromatin remodeling, as well as in viral infection, this makes the locus likely causal in SCZ, specifically through placental DNAm.

The placental DNAm sites that resulted from the SMR analyses confronting the BIP, MDD and SCZ GWAS, and the STB- and TB-imQTLs are also worthy of mention. For example, we found that cg27130493 is pleiotropically associated with BIP while it changes its methylation levels as a function of TB proportion, in a rs11637580 genotype-dependent manner. This CpG site is located in the gene body of *SMAD3,* a gene with singular functions in placenta, as well as in BIP itself[310,311]. Additionally, the SNP is not directly associated with BIP, increasing the likelihood of causality rather than pleiotropy. The most important limitation of this part of the study is the lack of placental cell type-specific expression data to ascertain whether cg27130493 and other DNAm sites identified in this approach are correlated with the expression of nearby genes in placenta, in a cell type specific manner. However, the fact that the CpG sites change not only depending on the genotype of adjacent SNPs, but also as a function of estimated placental cell proportions, increases the probability of the placenta being the effector organ of those genetic associations.

Lastly, we tried to ascertain the tissue-specificity of our findings by comparing placenta SMR hits to those from two brain *cis*-mQTL databases. We found a limited overlap among tissues, suggesting that the majority of our pleiotropically associated DNAm sites could be relatively specific of the placenta (or at least, not related to brain DNAm). However, it is important to consider that although the SMR approach was executed exactly in the same manner for the three databases, these were quite different from each other. In particular, the fetal brain database was calculated in a limited number of fetal brain samples, in SNP-CpG windows 10 times smaller than the ones we used (0.05 vs 0.5 Mb), and DNAm was measured with the Infinium HumanMethylation 450K array (with approximately half the probes in the Infinium HumanMethylation EPIC array used in the present study), therefore resulting in half a million *cis*-mQTLs compared to the more than 9 million

mQTLs that were included in the SMR analysis in our case. This suggests a more limited mapping potential with the fetal brain mQTL database and thus, a possible underestimation of the fetal brain DNAm that is really involved in the disorders studied.

The main limitation of our study is that, as pointed out by Ursini *et al*. in relation to their own work[211], the considerable overlap in cell biology between brain and placenta does not allow to exclude the possibility that part of the pleiotropy observed here is related to a more direct effect in the brain exerted by the same DNAm sites as observed in placenta. However, we believe that the different pieces of evidence, including the intersection with the placental eQTMs and even the presence of secondary signals, support that part of the genetic risk to suffer from BIP, MDD and very especially SCZ, could act through placental DNAm at specific genomic loci. Additionally, as it has been mentioned in the *Maternal ppBMI EWAS* discussion section, most of the samples were individuals from European descent, thus limiting the generalizability of our findings to other populations. Another limitation is that different methylation BeadChips have been employed in the INMA and RICHS cohorts, and thus, we are probably underestimating the effects of DNAm over placental gene expression. Finally, at this point we lack direct experimental support for our results. Further research, especially regarding *in vitro* validation of the most likely causal candidates, would enormously help supplement our findings.

In conclusion, we find placental *cis*-mQTLs to be highly placenta-specific, with a remarkable enrichment in genomic regions active in placenta and neurodevelopment- and mental health-related pathways. We prove that they are useful to map the etiologic window of neuropsychiatric disorders to prenatal stages and conclude that part of the genetic risk of BIP, MDD and in particular SCZ, act through placental DNAm at specific genomic loci. In fact, some of the observed associations might be causal rather than pleiotropic due to the presence of secondary association signals in conditional analyses, involvement of cell type-specific imQTLs and association with the expression levels of relevant genes in

placenta. It is of particular interest that SCZ-associated placental DNAm correlates with the expression of immune-related genes in placenta, providing further support to the hypothesis of the neurodevelopmental origins of SCZ. Future work, including functional approaches such as *in vitro* modulation of expression in potential effector cell lines, as well as long term follow up in prospective studies, will be needed in order to better characterize our findings and their implication in disease development.

FINAL REMARKS

Final remarks

On the one hand, we present here the largest placental epigenome-wide association meta-analysis carried out so far, with 2,631 mother-child pairs from European, North American and Australian PACE cohorts. The aim of this work was to study the effect of maternal ppBMI on placental DNAm. We identified 27 ppBMI-associated CpG sites, mainly located in open sea regions and completely depleted of CpG islands, with changes up to 2% lower DNAm per 10-unit higher ppBMI. Although such a difference in an individual woman is unlikely in the context of pre-pregnancy interventions, we consider that it could represent the difference between women in the normal range of BMI and women in the obesity category. Among the genes identified are *GPX1* and *LGR4,* both of which have been described in the context of obesity and cardiometabolic traits, highlighting the possibility that they may play a role in the link between maternal obesity, fetal growth and cardiometabolic outcomes in the offspring.

The CpG sites sensitive to maternal ppBMI presented biological similarities with cancer, probably due to the invasive abilities of both placental STBs and cancer. Additionally, the presence of oxidative-stress pathways reinforces previous evidence relating obesity-induced oxidative stress and the risk for adverse outcomes in pregnancy.

Another interesting finding is related to *ZNF217*, as this gene has been linked to adverse *in utero* conditions when it is epigenetically altered in placenta, and previously described with transcriptional regulation ability specifically in placenta. However, additional studies need to be done to elucidate whether it has an effect in epigenetic mechanisms. Similarly, we have observed an important overlap between the CpG sites identified and BW-associated regions, but further studies need to be performed to elucidate the suggested role of both fetal genetics and placental epigenetics in the regulation of fetal growth.

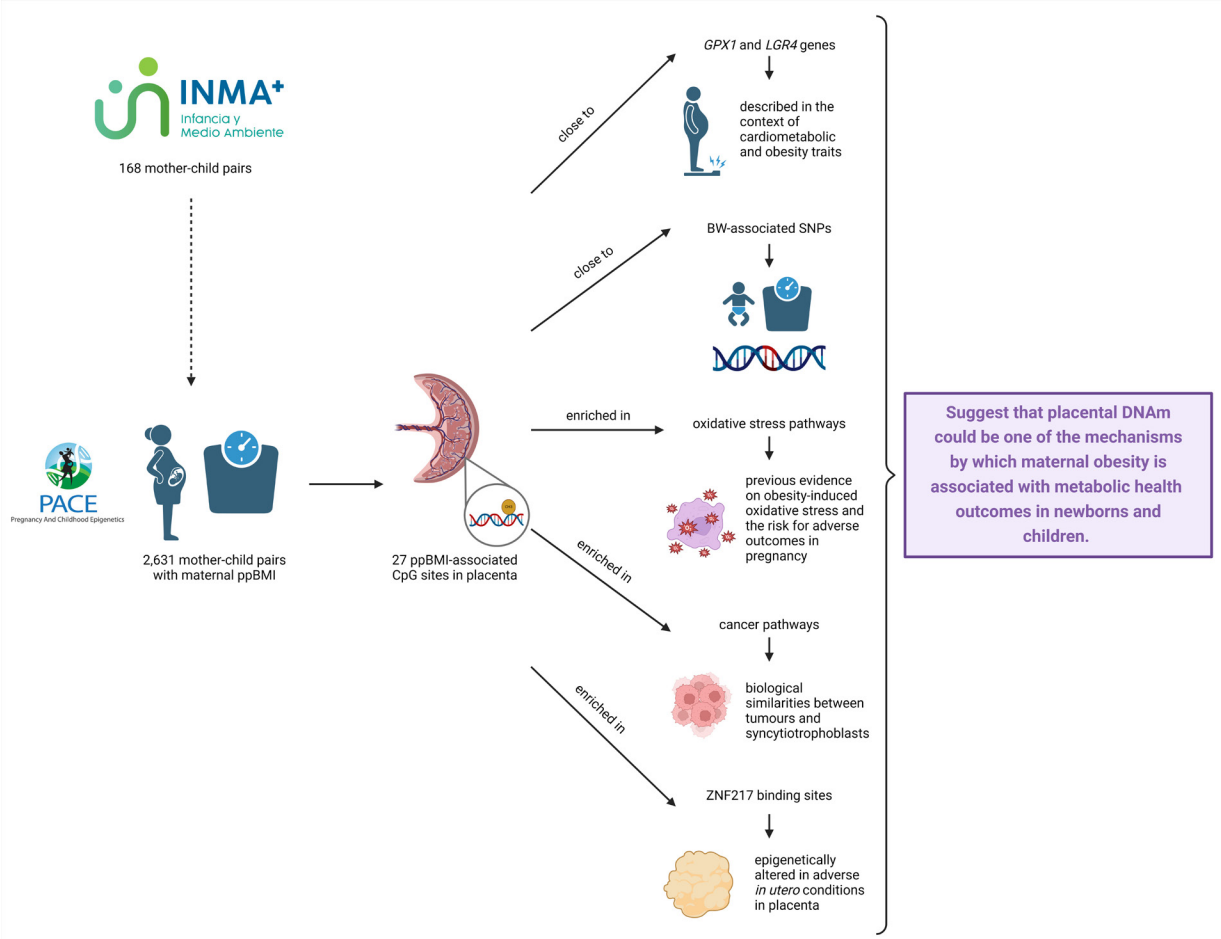


Figure 19. Schematic diagram of the final remarks of the maternal ppBMI EWAS meta-analysis.

On the other hand, we have also constructed several placental mQTL databases from 368 placenta samples from the INMA project. Importantly, we have made all results publicly available both in their raw formats and by means of a user-friendly, shiny-based browser that enables us to search for the mQTLs of interest by CpG, SNP and/or genomic coordinates. Briefly, we have identified 214,830 placental DNAm sites or CpG sites regulated by close genetic variants. These CpG sites were depleted in regions usually hypomethylated, highly likely because lower and more stable DNAm values will hardly correlate with the genotype or any other variable. In turn, an enrichment of intermediate DNAm values was observed, as well as an enrichment of placental

specific active chromatin marks. In addition, we also observed an enrichment of genes involved in neuropathy and mood disorders.

We have also demonstrated that cell type proportions affect placental DNAm and identified 38 and 1 placental DNAm sites modelled as a function of STB and TB proportions, respectively, in a genotype-dependent manner. We observed higher statistical power for the most abundant cell type and therefore, a larger amount of STB-imQTLs compared to TB-imQTLs. However, there was a remarkable sharing and, as expected, the allelic effects in the overlapping imQTLs were negatively correlated between the two cell types. As TBs are known to differentiate into STB throughout gestation, and STB content in term placenta is positively correlated with GA at birth, we wanted to know whether cell type-imQTLs are equivalent to GA-imQTLs. This was not the case, revealing that GA has an independent effect on placental DNAm other than the TB-to-STB differentiation. Additionally, no significant sex-imQTLs were obtained.

Given the neuropathy and mood-disorder pathways for which placental mQTLs appeared to be enriched of, we decided to conduct a multi-omics study of the potential placental contribution to the developmental origins of different neuropsychiatric disorders.

Ten neuropsychiatric disorders were considered, including ADHD, AGR, ASD, BIP, INT, MDD, OCD, PD, SA and SCZ. Remarkably, placental DNAm is pleiotropically associated with BIP, MDD, and in particular with SCZ, while it seems to barely associate with early onset conditions such as ADHD and ASD. These results could arise from the sample sizes and power constraints of the available GWAS, as well as from the higher heritability of certain disorders, including BIP and SCZ. But the results observed with the correlations with the number of SMR hits and the remarkable overlap between the three SCZ studies, suggests that our findings are guided not only by the strength and annotability of the genetic basis of the diseases studied, but rather by a genuine association with placental DNAm. In addition, it is well known that BIP, MDD and SCZ share a common genetic background and therefore, it is plausible that part of the genetic risk could act through common processes at the same developmental stages.

In the case of SCZ, the most interesting genes are *VPS37B* and histones *H2BC6* and *H3C4*. These three genes are the most relevant genes contributing to the enrichment in immune-related pathways, including HCMV and HIV infection, thus reinforcing the hypothesis that the fetal origins of this condition might come from the interaction between genetics and ELC, especially MIA.

In the context of BIP and MDD, the genes that are most likely to be causally involved in these neuropsychiatric disorders are *SMAD3* and *LRFN5*, respectively. Furthermore, in the case of *SMAD3,* the presence of a pleiotropic association between BIP and the DNAm levels of a CpG site regulated not only by the genotype, but also as a function of STB estimated placental cell type proportion, increases the probability of the placenta being the effector organ of those genetic associations.

Finally, although we cannot discard that some of our findings could be reflecting what is occurring in other tissues, we tried to demonstrate the unique role of placenta by comparing our findings with those from cord blood and (fetal) brain tissues. Particularly, the ppBMI-associated CpGs were compared to the results obtained from a similar EWAS in cord blood, while the BIP, MDD and SCZ SMR analyses performed with our placental mQTL database were reproduced with previously generated brain and fetal brain databases. In both cases, a minimal overlap was found, thus confirming an at least partial tissue-specificity of our findings.

In conclusion, early life exposures and risk genetics seem to have an impact on the future health of the offspring by modifying placental DNAm. The placental signatures described here have been demonstrated to be, at least partially, unique for this organ, reinforcing its role in the context of the DOHaD hypothesis. This work also highlights the importance of conducting previous PACE consortium in placenta, given the organ-specific signatures identified. Finally, we have pointed to several potentially causal genes that, according to different pieces of evidence, could underlie the pathogenesis of different neuropsychiatric disorders. Among others, the intersection of pleiotropy with placental eQTMs, the regional pleiotropy reported, and even the presence of secondary association signals derived from conditional analyses, strongly suggests that part of the genetic risk to suffer from BIP, MDD and very especially SCZ, acts

through placental DNAm at specific genomic loci. Future perspectives need to include studies with larger sample sizes and participants from other populations to increase statistical power and to stop the systematic exclusion of non-European descendants.

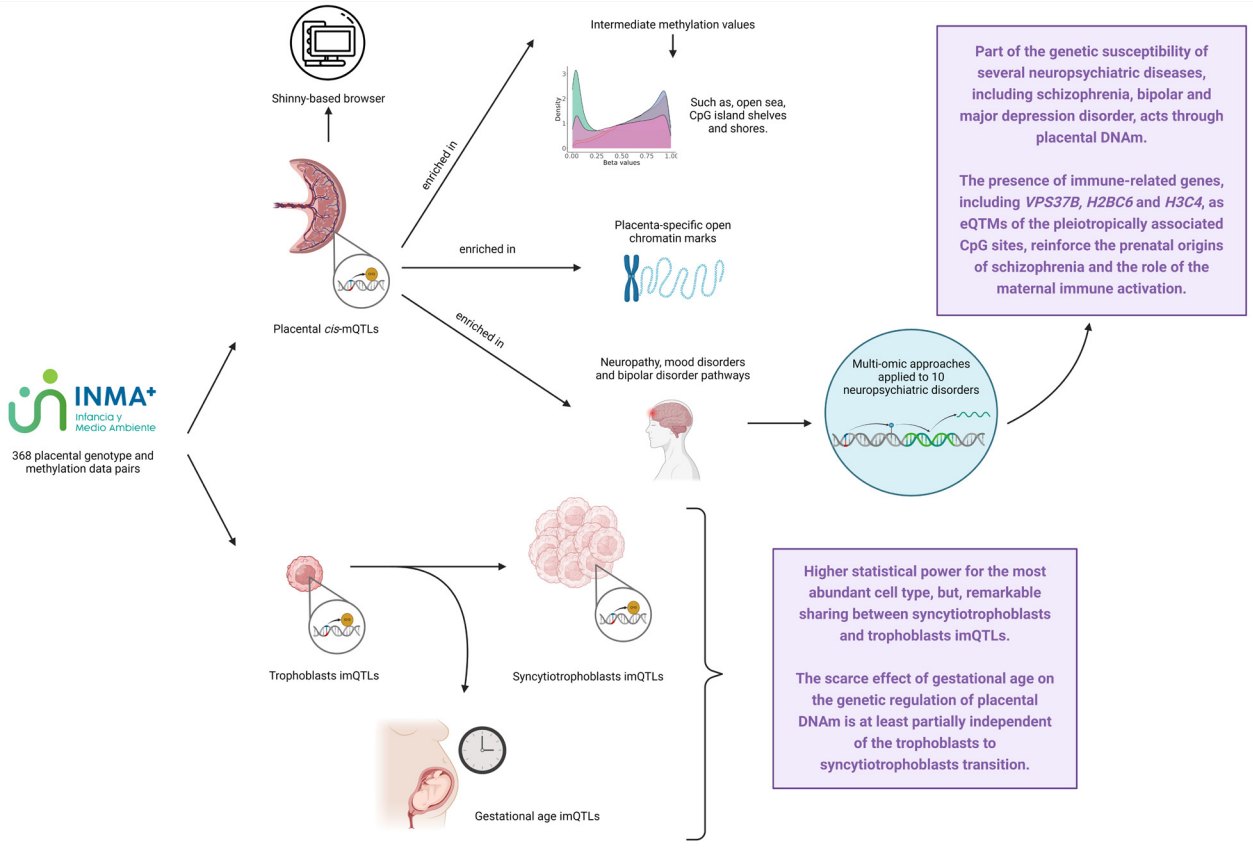


Figure 20. Schematic diagram of the final remarks of the placental *cis*-(i)mQTLs and downstream application to neuropsychiatric disorders.

CONCLUSIONS

Conclusions

The conclusions of the doctoral thesis are:

1. Maternal ppBMI impacts placental DNAm. In this thesis we have identified 27 placenta-specific DNAm sites associated to maternal ppBMI, many of which are located in open sea regions, close to obesity-related genes such as *GPX1* and *LGR4* and altogether, enriched in cancer and stress oxidative pathways. We propose that placental DNAm could be one of the mechanisms by which maternal obesity is associated with metabolic health outcomes in newborns and children.
2. We have made publicly available a catalog of placental mQTLs, including 214,830 unique DNAm sites modelled as a function of the genotype of the fetuses. Additionally, we have found that cell type proportions affect placental DNAm and have identified 38 and 1 placental DNAm sites modelled as a function of the proportions of STB and TB, respectively, in a genotype-dependent manner.
    a. The placental *cis*-mQTLs are depleted on usually hypomethylated regions, and in turn, enriched in regions with intermediate DNAm values, such as CpG islands shelves and shores. Moreover, the *cis*-mQTLs seem highly placenta-specific and, interestingly, are enriched in placenta-specific active chromatin marks, and in neuropathy and mood disorder-related pathways, among others.
    b. We observe higher statistical power for the most abundant cell type and therefore, a larger amount of STB-imQTLs compared to TB-imQTLs. However, there is a remarkable sharing and, as expected, the allelic effects in the overlapping imQTLs are negatively correlated between the two cell types. The moderate sharing signal between STB- and GA-imQTLs suggests that the scarce effect of GA on the genetic regulation of placental DNAm is at least partially independent of the TB-to-STB transition.

3. Part of the genetic susceptibility to suffer BIP, MDD and particularly SCZ, acts through placental DNAm. The potential causality of several of the observed associations is reinforced by secondary association signals identified in the conditional analysis, the involvement of cell type-imQTLs, and the correlation of the DNAm sites identified with the expression levels of relevant genes in placenta. Additionally, the presence of immune-related genes, including *VPS37B, H2BC6* and *H3C4*, as eQTMs of the pleiotropically associated CpG sites, reinforce the prenatal origins of SCZ and the role of the MIA in it. In conclusion, the genetic risk of several neuropsychiatric disorders could operate through placental DNAm and associated gene expression in placenta.

CODE AVAILABILITY AND LINKS OF INTEREST

Code availability and links of interest

The code for the genotype and DNAm data quality control for the placental *cis*-mQTLs, as well as for the TensorQTL nominal mapping is available in this GitHub repository link: https://github.com/ariadnacilleros/Cis-mQTL-mapping-protocol-for-methylome. The rest of the analyses regarding the mQTLs are available in this GitHub repository link: https://github.com/ariadnacilleros/Cilleros-PortetA.etal. Will Rayner's preparation Perl script from Mark McCarthy's Group is available in the following link: https://www.chg.ox.ac.uk/~wrayner/tools/. The Michigan Imputation Server website can be found here: https://imputationserver.sph.umich.edu/index.html#. The LocusZoom website tool used to plot the locuszoom plots can be found here: https://my.locuszoom.org/.

All the placental *cis*-mQTL databases are publicly available online in the following address: https://irlab.shinyapps.io/shiny_mqtl_placenta/. The appendices can be found here: https://addi.ehu.es/handle/10810/66967. And the INMA project website with all the information regarding this prospective study can be found here: http://www.proyectoinma.org/. The PACE consortium website with all the studies performed on behalf of it can be found here: https://www.niehs.nih.gov/research/atniehs/labs/epi/pi/genetics/pace.

FUNDING

Funding

This thesis was funded by grants from Instituto de Salud Carlos III (PI18/01142 and PI21/01491), co-funded by the European Union, the Spanish Ministry of Science (MCIN PID2019-106382RB-I00), and the Basque Departments of Health (GVSAN2018111086 and GVSAN2019111085) and Education (GIC21/144-IT1734-22).

APPENDIX

Appendix

Scan the QR code or click on the links on the following page
to have access to the appendices.

Appendix 1.

Distribution of maternal ppBMI, demographic variables, birth outcomes, and covariates, by cohort.

IQR = Inter Quartile Range.

Appendix 2.

Meta-analysis results for the association between maternal ppBMI and placental DNAm, adjusted for cellular heterogeneity; the data included are from AQUA, EARLI, EDEN, Gen3G, GENEIDA, HEBC, INMA, ITU, NHBCS, & RICHS.

Beta = Effect size; HetIsq = Heterogeneity I^2 Value; HetDf = Degrees of Freedom for Heterogeneity Analyses; HetPVal = Cochran's Q-test P value; Chr = Chromosome; Pos = Position (hg19); UCSC_RefGene_Name = Gene Annotations from the Illumina Annotation file; UCSC_RefGene_Accession = Gene Accession Number from the Illumina Annotation file; UCSC_RefGene_Group = Within Gene location from the Illumina Annotation file; SBE = Single Base Extension.

Appendix 3.

Meta-analysis results for the association between maternal ppBMI and placental DNAm, unadjusted for putative cellular heterogeneity; the data included are from AQUA, EARLI, EDEN, Gen3G, GENEIDA, HEBC, INMA, ITU, NHBCS, & RICHS.

Beta = Effect size; HetIsq = Heterogeneity I^2 Value; HetDf = Degrees of Freedom for Heterogeneity Analyses; HetPVal = Cochran's Q-test P value; Chr = Chromosome; Pos = Position (hg19); UCSC_RefGene_Name = Gene Annotations from the Illumina Annotation file; UCSC_RefGene_Accession = Gene Accession Number from the Illumina Annotation file; UCSC_RefGene_Group = Within Gene location from the Illumina Annotation file; SBE = Single Base Extension.

Appendix 4.

CpG sites from the leave-one-out analysis.

Source data underlying Figure 8. Cohort = cohort excluded in each analysis; Beta = Effect size; CI 95 lower = lower bound of the 95% confidence interval; CI 95 upper = upper bound of the 95% confidence interval.

Appendix 5.

CpG sites associated with maternal ppBMI in placenta and cord blood.

Beta = Effect size.

Appendix 6.

Contingency table of significant mSites from the permuted *cis*-mQTL database, and Relation to Island and UCSC RefGene annotation from IlluminaHumanMethylationEPICanno.ilm10b4.hg19 R package.

For this enrichment analysis the promoter region is defined from the 5' UTR to the 1st exon, including TSS1500 and TSS200 from the Illumina EPIC manifest.

Appendix 7.

List of the top 10,000 mSites with their corresponding top P values used for the eFORGE analysis (CpG list). eFORGE results from the top 10,000 mSites from the permuted database. Enrichment and depletion of overlap with tissue-specific DHS, H3 and H3K4me1 marks from consolidated the ROADMAP Epigenomics Mapping Consortium (DHS, H3 all marks, H3K4me1).

Zscore = Z score of the test data enrichment count versus the background; Pvalue = P value of the test data enrichment count versus the background; Cell = cell for which the enrichment is calculated; Tissue = tissue of the cell according either ENCODE or BioSamples; Datatype = Regulatory feature assessed; File = ROADMAP Epigenomics Mapping Consortium file; Probe = probes that overlap features in that file; Accession = GEO accession number of the data set; Qvalue = Q value of the test data enrichment versus the background.

Appendix 8.

Over-representation analysis of the permuted database genes performed with MissMethyl R package and GO and KEGG sets (GO, KEGG).

The genes used for this enrichment analysis were the ones annotated in the Illumina EPIC manifest, but in this case, MissMethyl takes in count the differing number of probes per gene present on the array, and the CpG sites that are annotated to multiple genes. ID = gene set ID from the database; Ontology (only in GO) = ontology that the GO term belongs to, "BP" - biological process, "CC" - cellular component, "MF" - molecular function; Term = gene set term; N = number of genes in the gene set; DE = number of genes that are differentially methylated; P.DE = P value of the over-representation analysis for the gene set term; FDR = P value adjusted by FDR.

GSEA of the permuted database genes performed with DOSE R package and DO sets (DO).

The genes used for this enrichment analysis were the ones annotated in the Illumina EPIC manifest. ID = gene set ID from the DO database; setSize = gene set size; enrichmentScore = Enrichment score; NES = Normalized enriched score; pvalue = P value of the gene set enrichment analysis; p.adjust = P value adjusted by Benjamini-Hochberg; qvalue = Q value of the genet set enrichment analysis; rank = Rank position of the gene set in the ranked gene set list; leading_edge = reports the tags that indicate the percentage of genes contributing to the enrichment score, the list which indicates where in the list the enrichment score is attained, and the signal that indicates the enrichment signal strength; core_enrichment = core enriched genes that contribute to the enrichment of GSEA analysis.

Appendix 9.

STB-imQTLs and TB-imQTLs mapped with TensorQTL (STB-imQTLs and TB-imQTLs).

The top mVariant is highlighted in blue. SNP = corresponds to the chromosome:position ID according to hg19; Tss_distance = Distance between the mSite site and the mVariant in bp; af = allele frequency of the tested allele; ma_samples = number of samples with at least one copy of the minor allele; ma_count = number of minor alleles across samples; pval_g = P value of the linear regression between the mSite methylation levels and the genotype of the mVariant; b_g = effect size of the linear regression between the mSite methylation levels and the genotype of the mVariant; b_g_se = standard error of the linear regression between the mSite methylation levels and the genotype of the mVariant; pval_i = P value of the linear regression between the mSite methylation levels and the interaction term; b_i = effect size of the linear regression between the mSite methylation levels and the interaction term; b_i_se = standard error of the linear regression between the mSite methylation levels and the interaction term; pval_gi = P value of the linear regression between the mSite methylation levels and the interaction between the interaction term, and the genotype of the mVariant; b_gi = effect size of the linear regression between the mSite methylation levels and the interaction between the interaction term, and the genotype of the mVariant; b_gi_se = standard error of the linear regression between the mSite methylation levels and the interaction between the interaction term, and the genotype of the mVariant; test_emt = effective number of independent variants in the *cis*-window estimated with eigenMT; pval_gi_x_ntest = P value of the linear regression between the mSite methylation levels and the interaction term, and the genotype of the mVariant, multiplied with the effective number of independent variants in the *cis*-window estimated with eigenMT.

TB-imQTLs and GA-imQTLs mapped with TensorQTL used to compute Storey pi1 (TB-imQTLs and GA-imQTLs Storey).

The description of the columns corresponds to the same as the previous STB-imQTLs and TB-imQTLs sheet.

Appendix 10.

SMR results from ADHD, ASD, BIP, MDD and SCZ. (ADHD, ASD, BIP, MDD and SCZ).

ProbeChr = CpG site chromosome; Gene = CpG site gene annotated in R package IlluminaHumanMethylationEPICanno.ilm10b4.hg19; Probe_bp = CpG site bp location; topSNP = CpG site top mVariant, most significant; topSNP_chr = top mVariant chromosome; topSNP_bp = top mVariant bp location; A1 = top mVariant tested allele; A2 = top mVariant other allele; Freq = top mVariant tested allele frequency; b_GWAS = effect size of the effect allele top mVariant in original GWAS; se_GWAS = standard error of the effect allele top mVariant in original GWAS; p_GWAS = P value of the effect allele top mVariant in original GWAS; b_eQTL = effect size of the mQTL in INMA nominal *cis*-mQTL database; se_eQTL = standard error of the mQTL in INMA nominal *cis*-mQTL database; p_eQTL = P value of the mQTL in INMA nominal *cis*-mQTL database; b_SMR = effect size from SMR test; se_SMR = standard error from SMR test; p_SMR = P value from SMR test; p_SMR_multi = P value from multi-SNP-based SMR test; p_HEIDI = P value from HEIDI test; nsnps_HEIDI = number of SNPs used in the HEIDI test; bonf = P value from multi-SNP-based SMR test, Bonferroni-adjusted.

Appendix 11.

Colocalization results from BIP, MDD and SCZ (BIP, MDD, SCZ).

In green, common hits (CpG sites) between SMR test and colocalization, and in yellow, common hits (CpG sites) between SMR test, colocalization and eQTMs. Region = original genomic region from the GWAS; Lead SNP = original lead SNP from the GWAS; nsnps = number of SNPs analysed; Position = base pair position of the CpG; PP.H0.abf = posterior probability of H_0 being true; PP.H1.abf = posterior probability of H_1 being true; PP.H2.abf = posterior probability of H_2 being true; PP.H3.abf = posterior probability of H_3 being true; PP.H4.abf = posterior probability of H_4 being true; PP.H3.abf+PP.H4.abf = sum of the posterior probabilities of H_3 and H_4 being true.

Appendix 12.

eQTMs results (FDR < 0.05) from the SMR significant CpG sites in MDD and SCZ. In yellow, common hits (CpG sites) between SMR test, colocalization and eQTMs.

Gene = Ensembl gene id; hgnc_symbol = HUGO Gene Nomenclature Committee (HGNC) gene symbol; chr = CpG site and gene chromosome; band = CpG site and gene cytogenetic band; beta = effect size of the linear regression between CpG methylation values and gene expression in placenta; t-stat = t-test statistic of the linear regression between CpG methylation values and gene expression in placenta; P value = P value of the linear regression between CpG methylation values and gene expression in placenta; FDR = P value adjusted by FDR; slope = slope value of the linear regression between CpG methylation values and gene expression in placenta; SCZ = CpG site SMR significant in SCZ; MDD = CpG site SMR significant in MDD.

Appendix 13.

Reactome results from the significant eQTM-genes from the SCZ analysis.

Entities found = number of analysed genes involved in the pathway; entities total = number of total genes involved in the pathway; entities ratio = the proportion of Reactome pathway genes represented by this pathway; entities pValue = the result of the statistical test for over-representation; entities FDR = P value adjusted by FDR; reactions found = the number of reactions in the pathway that are represented by at least one gene in the submitted data set; reactions total = the number of reactions in the pathway that contain genes; reaction ratio = the proportion of Reactome reactions represented by this pathway.

Appendix 14.

List of CpG sites considered for the conditional analysis failing HEIDI test in the first SMR analysis and with PP3+PP4 > 0.9 in Coloc (List of CpGs). Top mVariants conditioned in the nominal *cis*-mQTL database (Placental mQTLs db conditional).

Tss-distance = distance between mSite and mVariant in bp; nominal pvalue = P value of the conditioned mQTL; beta = effect size of the conditioned mQTL.

Top SNPs conditioned SCZ GWAS (SCZ GWAS conditional).

A1 = tested allele; A2 = other allele; freq = frequency of the tested allele; b = effect size of the tested allele; p = P value of the tested allele; n = sample size.

SMR results from the top SNPs conditioned in the SCZ GWAS and in the nominal *cis*-mQTL database (SMR conditional).

ProbeChr=CpGsitechromosome;Gene=CpGsitegeneannotatedinIlluminaHumanMethylationEPICanno.ilm10b4.hg19 R package; Probe_bp = CpG site bp location; topSNP = top mVariant, most significant; topSNP_chr = top mVariant chromosome; topSNP_bp = top mVariant bp location; A1 = top mVariant tested allele; A2 = top mVariant other allele; Freq = top mVariant tested allele frequency; b_GWAS = effect size of the effect allele top mVariant in conditional SCZ GWAS; se_GWAS = standard error of the effect allele top mVariant in conditional SCZ GWAS; p_GWAS = P value of the effect allele top mVariant in conditional SCZ GWAS; b_eQTL = effect size of the mQTL in SCZ conditional INMA nominal *cis*-mQTL database; se_eQTL = standard error of the mQTL in SCZ conditional INMA nominal *cis*-mQTL database; p_eQTL = P value of the mQTL in SCZ conditional INMA nominal *cis*-mQTL database; b_SMR = effect size from SMR test; se_SMR = standard error from SMR test; p_SMR = P value from SMR test; p_SMR_multi = P value from multi-SNP-based SMR test; p_HEIDI = P value from HEIDI test; nsnps_HEIDI = number of SNPs used in the HEIDI test; bonf = P value from multi-SNP-based SMR test, Bonferroni-adjusted.

Appendix 15.

STB-imQTLs and TB-imQTLs mapped with TensorQTL with nominal P value < 5x10-8 (STB-imQTLs, TB-imQTLs).

SNP = corresponds to the chromosome:position ID according hg19; Tss_distance = Distance between the mSite and the mVariant in bp; af = allele frequency of the tested allele; ma_samples = number of samples with at least one copy of the minor allele; ma_count = number of minor alleles across samples; pval_g = P value of the linear regression between the mSite methylation levels and the genotype of the mVariant; b_g = effect size of the linear regression between the mSite methylation levels and the genotype of the mVariant; b_g_se = standard error of the linear regression between the mSite methylation levels and the genotype of the mVariant; pval_i = P value of the linear regression between the mSite methylation levels and the interaction term; b_i = effect size of the linear regression between the mSite methylation levels and the interaction term; b_i_se = standard error of the linear regression between the mSite methylation levels and the interaction term; pval_gi = P value of the linear regression between the mSite methylation levels and the interaction between the interaction term, and the genotype of the mVariant; b_gi = effect size of the linear regression between the mSite methylation levels and the interaction between the interaction term, and the genotype of the mVariant; b_gi_se = standard error of the linear regression between the mSite methylation levels and the interaction between the interaction term, and the genotype of the mVariant; FDR = pval_gi adjusted by False

Discovery Rate; Class = classification of the interaction mQTLs according to the beta effects from the high and low percentiles as established in Kim-Hellmuth *et al.* 2022.

SMR results from BIP and MDD, and positive STB-imQTLs and TB-imQTLs (STB-imQTLs vs BIP, TB-imQTLs vs BIP, TB-imQTLs vs MDD).

ProbeChr=CpGsitechromosome;Gene=CpGsitegeneannotatedinIlluminaHumanMethylationEPICanno.ilm10b4.hg19 R package; Probe_bp = CpG site bp location; topSNP = CpG site top mVariant, most significant; topSNP_chr = top mVariant chromosome; topSNP_bp = top mVariant bp location; A1 = top mVariant tested allele; A2 = top mVariant other allele; Freq = top mVariant tested allele frequency; b_GWAS = effect size of the effect allele top mVariant in the original GWAS; se_GWAS = standard error of the effect allele top mVariant in the original GWAS; p_GWAS = P value of the effect allele top mVariant in the original GWAS; b_eQTL = effect size of the imQTL; se_eQTL = standard error of the imQTL; p_eQTL = P value of the imQTL; b_SMR = effect size from SMR test; se_SMR = standard error from SMR test; p_SMR = P value from SMR test; p_SMR_multi = P value from multi-SNP-based SMR test; p_HEIDI = P value from HEIDI test; nsnps_HEIDI = number of SNPs used in the HEIDI test; bonf = P value from multi-SNP-based SMR test, Bonferroni-adjusted.

Appendix 16.

SMR results from BIP, MDD and SCZ, and (fetal) brain *cis*-mQTLs (BIP brain, BIP fetal brain, MDD brain, MDD fetal brain, SCZ brain, SCZ fetal brain).

ProbeChr=CpGsitechromosome;Gene=CpGsitegeneannotatedinIlluminaHumanMethylationEPICanno.ilm10b4.hg19 R package; Probe_bp = CpG site bp location; topSNP = CpG site top mVariant, most significant; topSNP_chr = top mVariant chromosome; topSNP_bp = top mVariant bp location; A1 = top mVariant tested allele; A2 = top mVariant other allele; Freq = top mVariant tested allele frequency; b_GWAS = effect size of the effect allele top mVariant in the original GWAS; se_GWAS = standard error of the effect allele top mVariant in the original GWAS; p_GWAS = P value of the effect allele top mVariant in the original GWAS; b_eQTL = effect size of (fetal) brain mQTL; se_eQTL = standard error of (fetal) brain mQTL; p_eQTL = P value of (fetal) brain mQTL; b_SMR = effect size from SMR test; se_SMR = standard error from SMR test; p_SMR = P value from SMR test; p_SMR_multi = P value from multi-SNP-based SMR test; p_HEIDI = P value from HEIDI test; nsnps_HEIDI = number of SNPs used in the HEIDI test; bonf = P value from multi-SNP-based SMR test, Bonferroni-adjusted.

LABURPENA
RESUM
RESUMEN
SUMMARY

Sarrera

1. Osasunaren eta gaixotasunaren garapenaldiko jatorriaren hipotesia

Osasunaren eta gaixotasunaren garapenaldiko jatorriaren hipotesia (DOHaD, ingelesezko siglengatik) Barkerrek planteatu zuen 2007an. Hipotesiaren arabera, ingurune perinatalak eta bizitzako lehen urteek eragina izan dezakete osasun fetalean eta ondorengo bizitzan. Ingurumen-esposizio ugari aztertu dira, tartean amaren alkohola hartzea eta tabakismoa, bestelako osasunerako emaitza kaltegarriekin batera, hala nola nahasmendu neurologikoak edo kognitiboak eta gaixotasun kardiobaskularrak. Testuinguru horretan, ingurumen-erasoek umetokiaren ingurumena eralda dezaketela proposatu da, eta, ondorioz, bai fetuaren garapenari bai ondorengo osasunari eragiten diotela, plazentaren bidez.

1.1. Plazentaren papera DOHaDen

Plazenta da garatzen den lehen organoa, eta funtsezko zeregina betetzen du haurdunaldi osoan zehar, amaren eta fetuaren artean elikagaiak, gasak, hondakinak eta seinale endokrinoak elkartrukatzen baititu. Hori horrela, kokapen ezin hobea du jaio aurreko esposizioak ebaluatzeko DOHaD hipotesiaren testuinguruan. Gainera, erregulazio genomiko plazentarioak bizitzako lehen urteetako ezaugarrietan duen osasun-inpaktu sakona Bhattacharyak eta bere lankideek frogatu zuten 2022an. Bizitzan aurrerago ere luza daiteke eragin hori, ezaugarri konplexuen aurrekari etiologiko gisa, eta, ondorioz, garapena programatu dezake. Azken urteotan ikusi denez, plazentaren informazio *omikoa*, genetikak partzialki kontrolatutako datu epigenomikoak barne, oso baliagarria da jaio aurreko esposizioaren eta fetuaren edota bizitzako lehen urteetako osasunaren arteko harreman ugariak aztertzeko.

2. Epigenetika, genetikaren eta ingurumen-faktoreen arteko zubia

DNAren metilazioa (DNAm) gene-adierazpena erregulatzen duen prozesu epigenetikoa da, eta ingurumenaren eta genomaren arteko zubitzat hartu da, eta, aldi berean, ingurumen-faktoreen eta genetikaren kontrolpean dago.

Ikusi denez, DNAm-ak aldaketa handiak jasaten ditu, bai haurdunaldian zehar, bai *in utero* baldintza aldakorrei erantzuteko. Egungo ikerketa askoren arabera, ingurumen intrauterinoak funtzio plazentarioa aldaeraz dezake DNAm-aren bidez. Hortaz, garrantzitsua da kontuan hartzea plazentako DNAm-aren ezaugarri bereizgarria: ehun eta zelula mota gehienetan CpG gune gehientsuenak hipomettilatuta edo hipermetilatuta egoten dira, eta, ondorioz, DNAm-aren banaketa bimodala da; aldiz, DNAm plazentarioa banaketa trimodala du, partzialki metilatutako domeinu (PMD) eta tarteko metilazio-mailako CpG posizioak ugariak direlako.

Hala ere, DNA plazentarioaren berezitasunak ehun osoan ez ezik, plazentako zelula motetan ere ikusten dira. DNAm profil bereizgarrienak trofoblastoetan (TB) antzeman zituzten Yuanek eta bere lankideek 2021ean. TB-ak sinzitiotrofoblasto (STB) bilakatuz joango dira haurdunaldian zehar, eta bukaeran plazentako zelula motarik ugariena izan arte, zelula epitelialen geruza jarraitu, multinukleatu eta espezializatua osatuz, amaren eta fetuaren arteko elkarrekintza biomolekularra orkestratzen duena.

Oso ikerlan gutxik, akaso batek ere ez, aztertu dute zelula-mota bakoitzaren epigenomaren ekarpena haurdunaldiko esposizioekiko erantzunean edota hainbat osasun-ondoriotara bideratzeko joeran. Plazentaren garrantziak alde batetik, eta ehun osoan zein zelula-mota bakoitzean behatutako DNAm profil bereziek bestaldetik, espekulatzera garamatzate genoma-ingurumen elkarrekintza eta DNAm-an duen eragina oso bereziak direla plazentan eta ikerketa sakonagoa merezi dutela.

2.1. Epigenoma osoko asoziazio-azterketak

Epigenoma osoko asoziazio-azterketen (EWAS) helburua genoma osoaren aldakuntza epigenetikoa aztertzea da, batez ere CpG guneetako DNAm, intereseko fenotipoekin

modu esanguratsuan lotzen diren metilazio maila aldatuko posizioak (DMP) eta eremuak (DMR) antzemateko.

2.1.1. Haurdunaldiaren eta Haurtzaroaren Epigenetika partzuergoaren azterlanak

Azken urteotan, odol periferikoko zein zilbor-hesteko DNAm haurdunaldiko ingurumen-faktoreekiko sentikorra dela frogatu duten hainbat EWAS zuzendu ditu AEBko Osasun Institutu Nazionaleko Haurdunaldiaren eta Haurtzaroaren Epigenetika (PACE) partzuergoak. Hala ere, ikerlan gutxiagok arakatu dute esposizioek plazentako DNAm-an izan dezaketen eragina; gutxi horiek, ordea, ehun horretako DNAm-aren sinadura espezifikoa erakutsi dute.

Beste alde batetik, amaren haurdunaldi aurreko gorputz masa indizeak (ppBMI) interes berezikoa da; izan ere, hazkunde fetal aberrantearekin, makrosomiarekin eta jaioberrien morbilitatea areagotuarekin lotzen da amaren ppBMI altua, baita haurdunaldiaren konplikazioekin ere, hala nola pre-eklanpsia, diabetes gestazionala, haurdunaldiko hipertentsioa, erditze goiztiarra eta zesarea. Ondorengoen bizitzan zehar agertu daitezkeen bestelako osasun-ondorioekin ere erlazioa baduela ikusi da, obesitate-arrisku handiagoa eta haurren errendimendu kognitibo okerragoa, besteak beste.

Amaren ppBMI-k zilbor-hesteko odolean duen eragina aztertu zuen PACE partzuergoak 2017an eta DNAm-arekin lotura handia duela frogatu zuen. Hala ere, efektu epigenetiko esanguratsu asko apalak zirela ikusi zuten, eta ez zuten bidezidor biologiko aberasturik antzeman, jaioberriaren profil epigenetikoan eragin ditzaketen mekanismoei buruzko galderak erantzun gabe utziz. Testuinguru horretan, zilbor-hesteko odolean zein odol periferikoan izaten diren aldaketa epigenetikoak sakon ikertu diren arren, amaren ppBMI-k plazentako DNAm-an izan dezakeen eragina ez da sakonki esploratu, eta ehun horretan egindako azterketen emaitza esanguratsuak urriak dira, ziurrenik lagin-tamaina mugatuagaitik.

2.2. Metilazio-ezaugarri kuantitatiboen lokusak

Metilazio-ezaugarri kuantitatiboen lokusak (mQTL) nukleotido bakarreko polimorfismoak dira (SNP), zeinetan aldakuntza genetikoa CpG dinukleotido baten DNAm mailekin lotzen den. Oro har, QTLak aldaera genetikoen ondorio funtzionalak

ulertzeko erabili dira, horietako asko eskualde ez-kodetzaileetan daudelako eta ezaugarri edota gaixotasun ugarirekin lotzen direlako. Hainbat giza ehun osasuntsuren genoma osoko DNAm profila ezaugarritu zuten Oliva eta lankideek 2021ean, eta erakutsi zuten mQTL-ek ezaugarri ugariekiko lotura molekularrak ager ditzaketela eta eQTL-ren analisi bidez alboratutako gene hautagaiak zehazteko balio dutela.

Testuinguru horretan, eta aipatu dugun moduan, DNAm profilak ehun-espezifikoak ez ezik, zelula-motaren araberakoak ere badira. Gainera, zelula-motarekiko elkarrekintza-QTL-ak (iQTL) aberastuta daudela ehun- eta zelula-espezifikotasunean aurkitu zuten Kim-Hellmuthek eta lankideek, eta, hortaz, ehun osoetako QTL-ek baino bereizmen finagoa erdiesten dutela. Beraz, iQTLak oso tresna baliotsuak dira, ezaugarri konplexuekiko asoziazioen azalpen mekanistikora hurbiltzeko aukera ematen baitigute.

2.2.1. mQTL plazentarioen azterketak

Azken urteotan, mQTL plazentarioak mapatzeko hainbat ikerketa egin da (adibidez, Do *et al.* 2016, Delahaye *et al.* 2018, Tekola-Ayele *et al.* 2022 eta Casazza *et al.* 2024). Hala ere, oraindik ez dago sarbide publikoa duen mQTL plazentarioen datu-baserik, ez Illumina 450K matrizeko CpG posizioez besteko DNAm neurtzen duen analisirik, ezta ere lagin-tamaina handian egindakoa edo zelula-motaren heterogeneotasuna eta haien profil epigenetiko espezifikoak kontuan hartzen dituena.

3. Ausazkotze Mendeldar eta Kolokalizazio analisiak

Dagoeneko, genoma osoko asoziazio analisiek (GWAS) giza ezaugarri konplexuekin lotutako milaka aldaera genetiko identifikatu dituzte. Hala ere, aldaera horiek ezaugarrietan dituzten efektuen bitartekari diren geneak edo DNAren elementu funtzionalak ezezagunak dira oraindik. Fenotipoen arteko harreman kausalak inferitzea erronka garrantzitsua da eta berebiziko ondorioak ditu osasunaren edota gaixotasunen etiologia ulertzeko. Kolokalizazio eta Ausazkotze Mendeldar (MR) moduko analisiak erabilgarriak izan daitezke bi ezaugarrien arteko kausa-ondorioak estatistikoki inferitzeko, hau da, ezaugarri molekular baten (hots, DNAm) eta gaixotasun baten artekoak.

Lokus bereko bi asoziazio-seinale independenteak aldaera kausal berarekin bat datozen galdetzen dute kolokaliziazio analisiek. Erantzuna positiboa bada, ezaugarriak kolokalizatuta daudela esaten dugu, eta bi ezaugarri horiek kausazko mekanismoa partekatzearen probabilitatea nabarmen handitzen da. Mota honetako analisien ohiko adibidean eQTL analisia eta gaixotasunaren asoziazio-emaitzak erkatzen dira, gaixotasunaren gene kausala eta eragina dueneko ehuna zehazteko.

Bestalde batetik, aldakuntza genetikoa erabiltzen du MR-k esposizio moldagarriek askotariko ondorioetan eragin kausala aztertzeko. Mendelen herentzia-legeetan eta aldagai instrumentalak kalkulatzeko metodoetan oinarritzen da MR, eta kausa-ondorioak inferitzen ahalbidetzen dute, nahasmendu-faktoreen presentzian. Labur esanda, MR metodologia erabil daiteke SNP baten eta ondorio baten (gaixotasuna kasu) arteko harremanean DNAm bezalako esposizioa bitartekari ote den aztertzeko

4. Ezaugarri neuropsikiatrikoen eta jaio aurreko ingurumenaren arteko asoziazioa

Geroz eta argiago dago haurretan edo heldu gazteetan agertzen den hainbat asaldura neuropsikiatrikok jatorri fetala dutela. Horrela, amaren osasunak eta ingurumen-esposizio intrauterinoek eragina izan dezakete garun fetalaren garapenean, sarritan bide fisiologiko konbergenteen bidez. Ondorioz, hainbat nahasmendu komun garatzeko arrisku-faktore esanguratsu gisa deskribatu dira, tartean espektro autistaren nahasmendua (ASD), arreta- eta hiperaktibitate-defizitagatiko nahasmendua (ADHD), eskizofrenia (SCZ) eta nahasmendu bipolarra (BIP). Adibidez, testuinguru horretan aztertutako neuropsikiatriarekin lotutako bestelako ondorio batzuk portaerazkoak dira, hala nola, barneratzeko portaerak (INT) eta kanporatzekoak, depresio-nahasmendu larria (MDD), nahasmendu obsesibo-konpultsiboa (OCD) eta haurren portaera agresiboa (AGR) barne.

Laburbilduz, gero eta gehiago dira haurdunaldiko esposizioak, amaren ppBMI kasu, eta askotariko ezaugarri zein nahasmenduen garapena lotzen dituzten azterlanak, DOHaD hipotesiaren testuinguruan. Era berean, ikerketa epigenetikoek genetikaren eta ingurumen-faktoreen arteko zubi gisa funtzionatzen dutela erakutsi dute, baina azterketa gehienak odol-ehun periferikoan eta zilbor-hestekoan egin dira, haurdunaldi

arrakastatsu batean funtsezko eragile gisa plazentak izan dezakeen ekarpen potentziala alde batera utzita. Azken urteotan, ikerketa berriek frogatu dute organo horrek marka epigenetiko espezifiko eta bakarrak dituela, zilbor-hesteko odolean edo beste edozein ehunetan identifikatutakoez oso bestelakoak direnak. Patroi horiek izan daitezke jaio aurreko esposizioek eta sentikortasun-geneek beren efektuak ezartzeko mekanismoetako bat. Ildo horretan, zenbait ezaugarri neuropsikiatrikok jatorri prenatala dutela iradoki da, baina oso artikulu gutxik hartu du epigenoma plazentarioa genetikak nahasmendu horiek garatzen lagun dezakeen mekanismo potentzial gisa.

Helburuak

Tesi honen helburu orokorra da aztertzea, batetik, amaren ppBMI-k, ingurumen-faktore fetal garrantzitsua den heinean, eta bestetik, fetuaren genetikak, ondorengoaren bizitzan zeharreko osasunean izan dezaketen eragina, plazentako DNAm-aren bitartez.

Azterketaren helburu operatiboak honako hauek izan ziren:

1. Amaren ppBMI-ren eta epigenoma osoko DNAm plazentarioareren arteko balizko asoziazioa zehaztea, eta amaren obesitatea eta ondorengoen etorkizuneko osasun-ondorioak lotu dezaketen mekanismoak ulertzea.
2. Orain arteko mQTL plazentarioen datu-baserik handiena eraikitzea eta eskura jartzea, DNAm plazentarioak etorkizuneko bizitzako osasunean duen eragina ulertzeko tresna berritzaile gisa.
    a. Genetikak, eta, zehazkiago, SNP-ek erregulatutako plazentako DNAm guneak identifikatzea eta ezaugarritzea.
    b. SNP-en eta zelula-proportzioen, adin gestazionalaren (GA) eta sexu fetalaren arteko interakzioak erregulatutako plazentako DNAm guneak identifikatzea eta ezaugarritzea.
3. Arrisku-faktore genetikoek hainbat gaixotasun neuropsikiatriko garatzen laguntzen duten mekanismoa DNAm plazentarioa ote den zehaztea, besteak

beste, ADHD, AGR, ASD, BIP, INT, MDD, OCD, izu-nahasmendua (PD), suizidio-saiakera (SA) eta SCZ.

Metodoak

1. Plazentazko biopsiak eta DNA erauzketa

Haurtzaroa eta Ingurumena (INMA – *Infancia y Medio Ambiente*) egitasmoak populazioa oinarri duen Espainiako ama-haur kohorte-azterlana da, eta haurdunaldian eta lehen haurtzaroan zehar aireko, uretako zein dietako ingurumen-kutsatzaileek haurren hazkundean eta garapenean duten eragina aztertzea du helburu. INMA kohortean, Asturiasko, Gipuzkoako, Sabadelleko eta Valentziako 2.506 ama-feto bikote jarraitu ziren umeak jaio arte, eta 489 plazenta jaso ziren. Plazentaren barne-aldetik biopsiak hartu ziren, eta DNA erauzi zen.

2. Datu genotipikoen kalitate-kontrola

Genoma osoko 509.450 aldaera genetikoren genotipaketa 397 INMA laginetan egin zen, Illumina GSA BeadChip matrizea erabilita. Genotipoaren kalitate-kontrola PLINK 1.9 softwarearekin egin zen, ohiko gomendioak jarraituz. Genotipoak Michigan Inputazio Zerbitzarian inputatu ziren, Haplotype Reference Consortium (HRC) panela erabiliz, eta hurrengo ezaugarrietako bat zuten aldaerak baztertu ziren: 0,9tik beherako r2 inputazio-kalitatea, %5etik beherako alelo txikiaren maiztasuna (MAF), 0.05etik beherako Hardy-Weinberg orekaren P balioa edo bi alelo baino gehiago. Analisi honetan, metilazio-datuak zituzten laginak baino ez ziren aintzat hartu. Azken datuen multzoa 368 laginez eta 4.171.035 SNP-ez osatua zegoen.

3. DNAm datuen kalitate-kontrola

Ekoizlearen protokoloari jarraituz, 190 eta 397 plazenta aztertu ziren, Illumina 450K eta EPIC matrizeekin, ppBMI-ren EWAS-ean eta mQTL plazentarioaren datu-basean, hurrenez hurren. Labur esanda, DNAm datu gordinak normalizatu egin ziren eta kalitate baxuko laginak eta zunda eztabaidagarriak baztertu ziren. Amaren ppBMI-

ren kasuan,168 lagin eta 405,301 DNA zunda edo CpG gune gorde ziren azken datuen multzoan. Aldiz, mQTL plazentarioen kasuan, azken datuen multzoan 368 lagin eta 747,486 DNA zunda edo CpG gune gorde ziren.

Gainera, laginen heterogenotasun zelularra refFreeCellMix eta Planet R paketeekin zenbatetsi zen, amaren ppBMI-ren eta mQTL plazentarioen azterketetan, hurrenez hurren.

4. Amaren ppBMI-ren eta DNAm plazentarioaren EWAS

CpGko gune bakoitzeko DNAm plazentarioaren beta balioen eta amaren ppBMI-ren (aldagai jarraitu gisa) arteko asoziazioa aztertzeko erregresio lineal sendoak egin ziren. Ereduak doitu ziren amaren adinaren, erditze-kopuruaren, hezkuntza-mailaren eta haurdunaldian erretzaren arabera. Bi EWAS eredu kalkulatu ziren, bata RefFreeCellMix paketearen bitartez zenbatetsitako zelula-konposizioaren arabera doituta eta bestea doikuntzarik gabe.

Metaanalisia PACE partzuergoko hamar kohorteko (AQUA, EARLI, EDEN, Gen3G, GENEIDA, HEBC, INMA, ITU, MARBLES, NHBCS eta RICHS) 2,631 laginekin egin zen, eta horietan guztietan INMA kohortean erabilitako DNAm kalitatea kontrolerako protokolo berbera erabili zen.

Azkenik, ppBMI-rekiko sentiberak diren CpG guneak aztertu ziren, alde batetik aberaste funtzionala aurkitzeko, eta beste aldetik jaiotza-ondorioekin asoziatutako SNP-ekiko edota PACE partzuergoan aurretik burututako zilbor-hesteko odoleko azterketan identifikatutako CpG guneekiko hurbiltasuna zehazteko.

5. cis-mQTL plazentarioen analisia

Genotipoaren eta DNAm mailen balio normalizatuen arteko erregresio linealak egin ziren. Eredua fetuaren sexuaren, osagai nagusi (PC) genetiko eta ez genetikoen, zelula-konposizioaren arabera doitu zen. CpG gune bakoitzaren mQTL-en mapaketarako tartea bere posiziotik ±0.5 Mb-ra ezarri zen.

Ondoren, aberaste-azterketak egin genituen hainbat anotazio funtzionalekin, hala

nola kromatina markak eta histona ehun-espezifikoak, bai eta Gene Ontology eta Disease Ontology datu-baseetako gene-multzoak ere.

6. Interakzioko *cis*-mQTL plazentarioen analisia

Zelula-motarekin elkarreragiten duten mQTLak (imQTLak, ingelesezko *interacting mQTLs* sigletatik) kalkulatu ziren STB eta TB zelula-motetan; aldi bakoitzean genotipoa eta zelula-mota interakzio-termino gisa hartzen zituzten eredu linealak eraiki ziren, eta fetuaren sexuaren eta DNAm-aren PC genetikoen eta ez-genetikoen arabera doitu. Horretaz gain, fetuaren sexuarekiko eta GA-rekiko imQTL-ak (GA-imQTL-ak) ere kalkulatu ziren, eta aldagai bertsuen eta zenbatetsitako heterogenotasun zelularraren arabera doitu ziren. Sexu-imQTL-en kasuan, fetuaren sexua doikuntza-aldagaikideetatik baztertu zen.

7. Arazo neuropsikiatriko batzuen balizko jatorri plazentarioa argitzeko ikuspegi multiomikoak

Lan honetan erabilitako GWAS-en laburpen-estatistikoak eskura dauden ADHD, AGR, ASD, BIP, INT, MDD, OCD, PD, SA eta SCZ ikerketarik handienetakoak izan ziren. Estatistiko horiek harmonizatu egin ziren dbSNPren 155. bertsioarekin eta INMA laginen genotipo-datuekin, erreferentzia gisa.

MR analisiak egiteko, CpG bakoitzaren DNAm esposiziotzat eta nahasmendu neuropsikiatrikoak ondoriotzat hartu ziren, eta horrela egiaztatu zen plazentako DNAm-ak ezaugarri horietan duen eragin kausala. Gainera, HEIDI izeneko heterogeneotasun-proba ere aplikatu zen; horren bitartez, lotura-desoreka indartsuan leudekeen bi SNP-ek esposizioari eta ondorioari modu independentean eragiteko aukera aztertu zen. MR azterketa bera aplikatu zen imQTL datu-baseetan, baina TB- eta STB-imQTLen kasuan, soilik hartu ziren kontuan zelula-mota bakoitzaren ugaritasunaren arabera eragin genotipiko nagusia handitua zutenak.

Kokatze-azterketa jatorrizko GWAS-ek definitutako lokus esanguratsuen eta plazentako mQTL-en artean egin zen. Azterketa hori xehetasun handiagoz aztertutako hiru fenotipoekin soilik egin zen (BIP, MDD eta SCZ).

Azkenik, identifikatutako CpG guneek plazentako adierazpen genikoan duten eragina ebaluatu dugu, RICHS kohorte iparramerikarreko adierazpen-ezaugarri kuantitatiboen metilazioen datu-base batean (eQTM, edo *expression quantitative trait methylation*, ingelesez). SCZ-riekin lotutako eQTM gene-multzoen aberaste-azterketak, Reactome datu-basea eta webguneko tresna erabili ziren.

8. *cis*-QTL plazenarioen eta nahasmendu neuropsikiatrikoen analisi baldintzatua

Goreneko mQTL-rako baldintzatuako analisia egin genuen MR analisia gainditu baina heterogenotasun-proban porrot egin zuten, zenbait kasutan heterogenotasun hori egiazko seinale sekundarioek bultzatu dezaketelako. Baldintzatutako azterketa xehetasun handiagoz aztertutako hiru fenotipoekin bakarrik egin zen (BIP, MDD eta SCZ).

Emaitzak

1. Amaren ppBMI 27 plazentako DNAm guneekin asoziatuta dago

Bonferroni proba anizkoitzetarako zuzenketa aplikatu ondoren, 27 eta 42 CpG gune identifikatu genituen, non amaren ppBMI modu esanguratsuan asoziaturik zegoen plazentako DNAm mailekin, zelula-motako osagaien arabera doitutako eta doitu gabeko ereduetan, hurrenez hurren. Zelula-osagaien arabera doitutako ereduan, amaren ppBMI altua DNAm plazentario baxuarekin zegoen asoziatuta 24/27 CpG gunetan, eta doitu gabeko ereduan, aldiz, 33/42 gunek agertu zuten asoziazioa. Nolanahi ere, metilazio diferentziala erakutsi zuten CpG guneen efektuaren tamainek ereduen arteko korrelazio positiboa zuten. Azkenik, kohorteen arteko asoziazioen heterogeneotasuna txikiagoa izan zen zelula-motaren proportzioen arabera doitutako ereduan, doitu gabeko ereduarekin alderatuta, eta, horregatik, erabat doitutako ereduaren emaitzekin jarraitu genuen, ondorengo analisiak egiteko.

Zelula-motaren arabera doitutako EWAS-eko 27 DMP-en artean, aipatzekoak dira CpG gune baka batzuk: cg08219219 CpG gunerik esanguratsuenean, adibidez, amaren ppBMI-ren 10 unitateko hazkundea CpG-aren metilazio-mailaren %1.1 baxuagoarekin lotzen da. Eragin positiborik handiena cg14704941 gunean antzeman zen, ppBMI-ren 10 unitateko igoerak plazentako DNAm-aren %2ko emendioa ekarri bait zuen. Era berean, eragin negatiboarik handiena cg04724807 CpG-an behatu zen, ppBMI-ren 10 unitateko DNAm-ak %1.8 egiten baitu behera.

Amaren ppBMI-rekiko sentiberak diren 27 CpG guneei anotatutako 26 geneak aberastuta zeuden zelula txikien birikako minbizian eta estres oxidatiboak induzitutako seinaleztapen-bideko gene-multzoetan. Horretaz gain, geneen zerrenda bera aberastuta zegoen *ZNF217* transkripzio-faktoreak erregulatutako genetan.

Era berean, aztertu genuen, alde batetik, amaren ppBMI-rekiko sentiberak CpG guneak jaiotza-ondorioekin asoziatutako SNP-etatik gertu ote zeuden, eta beste aldetik, amaren ppBMI-ren erlazionatutako plazentako DNAm-aren sinadura bat ote zetorren PACE partzuergoak zilbor-hesteko odoleean aurkitutako asoziazioekin. Alde batetik, ppBMI-rekin asoziatutako 27 CpG-ren herena baino gehiago jaiotza-pisuarekin lotutako SNP-en ±0.5 Mb-ko tartearen barruan zeuden. Bestetik, ez zen aurkitu amaren ppBMI-ri lotutako CpG komunik bi ehunen artean.

2. Plazentako 200,000 CpG gunetik gora *cis*-aldaera genetikoek erregulatutakoak dira

Guztira, 214,830 mQTL plazentario aurkitu ziren, 214,830 mSite (CpG gune) eta 150,649 mVariant (SNP) oinarri hartuta. Gehienetan, mVariant eta mSite partaideak elkarrengandik hurbil zeuden, genetikak modulatutako DNAm guneak aldaera erregulatzaileetatik gertu egon ohi direnaren seinale.

Oro har, mSite-ak urriak ziren CpG uharteetan zein promotoreetan, eta aldiz, aberastuta zeuden gorputz genikoetan eta tarteko metilazio-balioak izan ohi dituzten eremu genomikoetan, hala nola itsaso irekian eta CpG uharteen plataformako eta urertzeko eskualdeetan. Horrez gain, aberastuta zeuden bai giza antigeno leukozitarioaren (HLA) eskualdea bai plazentaren espezifikoak diren kromatina irekiko eta transkripzio

aktiboko eskualdeak. mSite-etatik gertuko geneak 219 gene multzotan aberastuta zeuden, neuropatia, gogo-aldartearen nahasmendua, gaixotasun desmielinizatzailea eta BIP barne.

mQTL plazentarioaren datu-basea helbide honetan aurki daiteke: https://irlab.shinyapps.io/shiny_mqtl_placenta/.

3. Zelula-motarekin elkarreragiten duten mQTL-ak detektatzen dira STB eta TB-etan.

Zelula-motarekiko interakzioari dagokionez, 38 STB-imQTL eta TB-imQTL bakarra aurkitu ziren, eta, aldiz, ez zen GA- edo sexu-iQTL esanguratsurik identifikatu. STB-imQTL kopurua handiena izateak zelula mota ugarienaren potentzia estatistiko handiagoa agerian uzten du. STB- eta TB-imQTL-en arteko partekatzea nabarmena zen, GA-ren eragin apalarekin. Horrek iradokitzen du TB-STB trantsizioak eragin handia duela plazentako metiloman, eta aldiz GA-ren eragin apala independentea dela, zati batean behintzat.

4. Hurbilketa multiomikoek zenbait nahasmendu neuropsikiatrikoren balizko jatorri plazentarioa argitzen dute

MR estrategiak ez zuen emaitza esangarrik lortu, OCD eta SA nahasmenduetan. Beste gaixotasunetan, ordea, hainbat DNAm mSite identifikatu zen: ADHD (n = 1), ASD (n = 1), BIP (n = 30), MDD (n = 28) eta, bereziki, SCZ (n = 214). MR analisietan identifikatutako asoziazio pleiotropikoak erreplikatzeko, kolokalizazio-analisiak egin ziren BIP, MDD eta SCZ gaixotasunetako GWAS-ek definitutako 64, 102 eta 287 eskualde genomikotan, hurrenez hurren. Bestalde, mQTL-ei dagokienez, FDR < 0.05 izan zuten 214,830 mSite-etatik, 12,228, 10,343 eta 38,412 mSite aukeratu genituen BIP, MDD eta SCZ gaixotasunetan, hurrenez hurren. Horrela, BIP-aren kasuan, 7,123 eskualde-CpG parek, 6,487 DNAm gunek eta 64 eskualdek agertu zuten kolokalizazio zantzua. MDD-an, aldiz, 3,895 izan ziren kolokalizazioa iradokitzen zuten eskualde-CpG pareak, 3,850 DNA guneak eta 79 eskualdeak. Azkenik, SCZ-ren kasuan, 21,071 eskualde-CpG parek, 19,661 DNA gunek eta 274 eskualdek kolokalizatu

zuten. MR analisien emaitzekiko gainezarpena aztertuta, BIP-aren kasuan, SMR-k identifikatutako 30 DNAm guneetatik, 28tan (%93) aurkitu zen kolokalizazio ebidentzia; MDD-ren kasuan, SMR-ren 28 emaitzetatik 20k (%71) erakutsi zuten kolokalizazio seinalea, eta SCZ-ren 214 SMR emaitzetatik, 198 DNAm gunek (%92) kolokalizatu zuten SCZrekin.

Ondoren, RICHS kohortean kalkulatutako eQTM-ak aztertu ziren. MDD eta SCZ gaixotasunetan SMR-k identifikatutako 28 eta 214 DNAm gune esanguratsuetatik, 3 eta 43 korrelazionatuta zeuden 4 eta 26 eQTM-generen adierazpen mailarekin, hurrenez hurren, eta emaitzek CpG gune horien funtzio plazentarioa azpimarratzen dute.

Jarraian, SMR, kolokalizazioa eta eQTM analisien emaitzen arteko gainezarpena identifikatu genuen hiru ezaugarrietan. MDD-ren kasuan, plazentako DNAm gune bat kolokalizatuta ageri zen 14. kromosomako asoziazio-lokus genomiko batekin, eta, gainera, CpG horren DNAm mailak *LRFN5* genearen adierazpen mailak korrelazionatuta zeuden. *LRFN5* TB zelula ametan adierazten de eta litekeena da funtzio garrantzitsua betetzea bertan. SCZ-ren kasuan, 40 plazentako DNAm gune kolokalizatuta zeuden SCZ-rekin asoziatutako 12 lokusekin (1., 3., 6., 7., 8., 11., 12., 18., 19. eta 22. kromosometan), eta DNAm eta adierazpen genikoaren arteko korrelazioa gertuko 26 eQTM-geneetatik 23 hauetan ikusi zen: *GLB1L3, H2BC6, H3C4, KCNG2, LETM2, NAGA, PGBD1, PSMG3, RNF39, SDCCAG8, SFMBT1, SLC6A16, TOB2P1, TRIM27, TXNL4A, VARS2, VPS37B, ZKSCAN4, ZNF165, ZNRD1-AS1, ZSCAN12, ZSCAN12P1* eta *ZSCAN23*.

Gene horietako batzuk aipatzekoak dira, *H2BC6* eta *H3C4* histonak, esaterako. Alde batetik, *H2BC6* abortu espontaneo errepikakorrak izan dituzten amen plazentan identifikatu da. Bestaldetik, *H3C4* desregulatuta dago neurulazio goiztiarrean zehar alkoholaren esposizioa jasan duten sagu-enbrioietan, DNAm patroien aldaketa aberranteak medio. Azkenik, *VPS37B* gene erabakigarria da kimatze biralean. Oro har, gene horiek homeostasi immunitarioan eta infekzio biralean parte hartzen dute, eta haurdunaldia garatzeko garrantzitsuak dirudite.

Azkenik, SMR eta eQTM analisien emaitzen gainezarpeneko 26 geneen aberaste-analisia egin genuen Reactome erabilita. Aberastutako 101 gene-multzo antzeman genituen, tartean erregulazio epigenetikoko bide ugari, eta horrek iradokitzen du

SCZren arrisku-aleloek plazentaren paisaia epigenetikoa alda dezaketela, besteak beste, lehen aipatutako histona genearen erregulazioaren bitartez. Gainera, aberaste nabarmena ikusten dugu sistema immunitarioan inplikatutako bideetan, hala nola giza zitomegalobirus infekzioa eta, oro har, infekzio birala. Horrek guztiak indartu egiten du amaren aktibazio immunitarioa (MIA) plazentaren eta SCZ-ren arriskuaren arteko lotura gisa proposatu duen hipotesia.

5. Balizko seinale kausal sekundarioa antzeman da SCZ-n

Baldintzatutako azterketaren bidez, bai cg27314558 posizioan bai SCZ gaixotasunean eragin pleiotropikoa zuen bi SNP independente barnebiltzen zuen eremua identifikatu genuen. CpG gune hau transferentziarako RNA (tRNA) ugariz inguratuta, gertuena alanina-tRNA izaki. Jakina denez, tRNA molekulen transferentzia plazentarioaren urritasuna lotuta dago hazkunde fetalaren murrizketarekin.

6. BIP eta SCZ eta STB- eta TB-imQTL-en arteko asoziazio pleiotropikoa

Zelula motaren imQTL-en eta BIP, MDD zein SCZ gaixotasunen arteko MR analisiek asoziazio pleiotropikoak antzeman zituzten BIP-arekin (STB-imQTL eta TB-imQTL bana) eta MDD-rekin (TB-imQTL). Esaterako, BIP-arekin asoziatutako TB-imQTL-an inplikatutako CpG gunea *SMAD3* genean dago. Frogatu dago gene horren plazentako gainadierazpenak bultzatzen duela TB-ek endotelioaren antzeko sareak sortzeko gaitasuna, eta pre-eklanpsiarekin erlazionatuta dago dela. Gainera, gene hau BIP-ean erabilitako litio tratamenduaren itua da.

Eztabaida

1. Amaren ppBMI EWAS

Lan honetan, inoiz egin den DNAm plazentarioaren EWAS handiena aurkeztu dugu. Bertan, amaren ppBMI-rekin asoziatutako 27 CpG gune identifikatu ditugu, eta

ppBMI-ren 10 unitateko alderako %0.5-2.0 arteko aldaketak ikusten ditugu DNAm. Gainera, aurkitutako DNAm aldaketa horiek plazentaren espezifikoak dirudite, gainezarpen txikia erakusten baitute aurretik zilbor-hesteko odolean egindako ppBMI-ren aztarketen meta-analisiarekin. Metilazio maila aldatuko CpG guneak itsaso zabaleko eskualdeetan daude batez ere, eta ez dira CpG uharteetan ageri; gainera minbiziarekin eta estres oxidatiboarekin zerikusia duten bidezidorretan aberastuta daude.

Horrekin guztiarekin, eta amaren ppBMI-rekin asoziatutako CpG-ak obesitatearekin lotutako geneetatik hurbil daudela ikusita, hurrengo hipotesi hau planteatu dezakegu: baliteke DNAm plazentarioa izatea amaren obesitateak hazkunde fetal aberrantea eragiteko mekanismoetako bat, eta, agian, ondorengoen etorkizuneko bizitzan ager daitezkeen osasun metabolikoaren bestelako ondorioak eragitekoa ere. Hala ere, ezin bazter dezakegu hautemandako metilazio-aldaketak ppBMI altuaren eraginpean egotearen markatzaileak besterik ez izatea, etorkizuneko osasunean ondorio kausalik ekarri gabe, eta, horrexegatik, haurdunaldiaren konplikazioetan edo epe luzerako ondorio metabolikoetan benetan paperik ote duten ulertzeko, beharrezkoa izango da gure aurkikuntzak azterketa funtzionalen edo inferentzia kausalaren analisien bitartez osatzea.

2. mQTL plazentarioak

Laburbilduz, aurkitu dugu mQTL plazentarioak oso plazentaren espezifikoak direla, eta aberaste nabarmena dutela plazentaren eskualde genomiko aktiboetan eta neurogarapenarekin zein osasun mentalarekin zerikusia duten bidezidorretan. Nahasmendu neuropsikiatrikoen leiho etiologikoa jaio aurreko aldira mapatzeko baliagarriak direla ere frogatu dugu, eta ondoriozta dezakegu BIP, MDD eta, bereziki, SCZ pairatzeko arrisku genetikoaren parte batek behintzat, lokus zehatzetako DNAm plazentarioaren bidez jarduten duela. Izan ere, badaude zantzuak behatutako asoziazio batzuk pleiotropikoak baino kausalak izan daitezkeela esateko, hala nola baldintzatutako analisietako seinale sekundarioak, zelula-motaren imQTL espezifikoen parte-hartzea eta plazentaren gene garrantzitsuen adierazpen-mailekiko korrelazioa. Interes berezia du SCZ-ari lotutako DNAm plazentarioa plazentaren

sistema immunologikoarekin zerikusia duten geneen adierazpenarekin bat etortzea, horrek babes gehigarria ematen baitio SCZ-aren jatorria neurogarapenean kokatzen duen hipotesiari. Etorkizunean, bestelako lanak beharko dira gure aurkikuntzak eta gaixotasunaren garapenean duten inplikazioa hobeto ezaugarritzeko, tartean hurbilketa funtzionalak, hala nola zelula-lerroen adierazpen genetikoaren modulazioa *in vitro*, edota epe luzeko jarraipena azterketa prospektiboetan,

Azken oharrak

Esposizio goiztiarrek eta arrisku genetikoak ondorengoen etorkizuneko osasunean eragina dutela dirudi, DNAm plazentarioaren aldaketaren bidez. Frogatu denez, hemen deskribatutako sinadura plazentarioak, zati batean behintzat, organo horren esklusiboak direla, eta horrek DOHaD hipotesiaren testuinguruan plazentak duen papera indartzen du. Lan honek agerian uzten du, halaber, garrantzitsua dela PACE partzuergoan plazentako azterketak aurretik egin izana, identifikatutako organoen berariazko sinadurak direla eta. Azkenik, kausazkoak izan daitezkeen zenbait gene aipatu ditugu, ebidentzia ezberdinen arabera, hainbat arazo neuropsikiatrikoren patogenesiaren azpian egon daitezkeenak, haien efektuak plazentan gauzatzean. Besteak beste, pleiotropia horizontalaren eta eQTM plazenarioen arteko gainezarpenak, aurkitutako eskualde-pleiotropiak, eta baldintzatutako analisietan aurkitutako seinale sekundarioak egoteak, argi iradokitzen dute BIP, MDD eta bereziki SCZ pairatzeko arrisku genetikoaren zati batek eragiten duela lokus genomiko espezifikoetako DNAm plazentarioaren bidez. Etorkizuneko ikuspegiek izan beharko dituzte europarraz besteko populazioetako lagin-tamaina handiagoko eta parte-hartzaile gehiagoko azterlanak, potentzia estatistikoa handitzeko eta Europakoak ez diren taldeen bazterketa sistematikoari amaiera emateko.

Ondorioak

Doktorego tesiaren ondorioak honako hauek dira:

1. Amaren ppBMI-k DNAm plazentarioari eragiten dio. Tesi honetan, amaren ppBMI-rekin lotutako 27 DNAm gune plazenta-espezifikoak identifikatu ditugu, eta horietako asko itsaso zabaleko eskualdeetan daude, obesitatearekin zerikusia duten geneetatik gertu, hala nola *GPX1* eta *LGR4*, eta, oro har, minbiziaren eta estres oxidatiboaren bidezidorretan aberastuta daude. Proposatzen dugu DNA plazentarioa izan daitekeela amaren obesitatea jaioberrien eta haurren osasun metabolikoaren ondorioekin lotzeko mekanismoetako bat.
2. Jendearen eskura mQTL plazenarioen katalogo bat jarri dugu, fetuen genotipoaren arabera modelatutako 214,830 DNAm gune bakar biltzen dituena. Gainera, zelula-proportzioek DNAm plazentarioari eragiten diotela aurkitu dugu, eta STB eta TB proportzioen arabera modelatutako DNAm plazentarioko 38 eta 1 genotipoaren menpeko DNAm gune identifikatu ditugu, hurrenez hurren.
    a. Plazentako *cis*-mQTL-ak normalean hipometilatutako eskualdeetan falta dira, eta, aldi berean, DNAm bitarteko balioak dituzten eskualdeetan aberastuta daude, hala nola CpG uharteetako plataformetan eta urertzetan. Gainera, *cis*-mQTL-ak plazentaren oso espezifikoak direla dirudi, eta, bitxia bada ere, plazentaren kromatina aktiboko marka espezifikoetan eta neuropatiekin eta gogo-aldartearen nahasmenduekin zerikusia duten bideetan aberastuta daude, besteak beste.
    b. Zelula-mota ugarienaren azterketarako potentzia estatistiko handiagoa ikusten dugu, eta, ondorioz, STB-imQTL kopuru handiagoa aurkitu dugu TB-imQTL-ekin alderatuta. Hala ere, partekatzea nabarmena da eta, espero zen bezala, gaiezarritako imQTL-etako aleloen efektuak modu negatiboan korrelazionatuta daude bi zelula mota biren artean. STB- eta GA-imQTL-en arteko partekatze

moderatuak iradokitzen du DNAm plazentarioaren erregulazio genetikoan GA-k duen eragin txikiaren parte bat, behintzat, TB-STB trantsizioarekiko independentea dela.

6. BIP, MDD eta, bereziki, SCZ garatzeko joera genetikoaren zati batek DNAm plazentarioaren bidez jarduten du. Badira behatutako zenbait asoziazioren kausalitate potentziala indartzen duten emaitzak, hala nola baldintzatutako azterketetan identifikatutako asoziazio seinale sekumdarioak, zelula-motako imQTLen parte-hartzea eta plazentako gene garrantzitsuen adierazpenaren eta identifikatutako DNAm guneen arteko korrelazioak. Gainera, sistema immunitarioarekin lotutako geneen presentziak, tartean asoziazio pleiotropikoko CpG guneen eQTM diren *VPS37B, H2BC6* eta *H3C4*, SCZ-ren jaio aurreko jatorria eta MIA-ren papera indartzen ditu. Laburbilduz, hainbat nahasmendu neuropsikiatrikoren arrisku genetikoak DNAm plazentarioaren bidez eta horrekin lotutako geneen plazentako adierazpenaren bidez eragin lezake.

Introducció

1. La hipòtesis sobre l'origen del desenvolupament de la salut i de la malaltia

La hipòtesi dels orígens del desenvolupament de la salut i la malaltia (DOHaD, per les seves sigles en anglès *Developmental Origins of Health and Disease*) va ser encunyada per Barker el 2007, i proposa que l'entorn perinatal i primerenc pot afectar la salut fetal i posterior. S'ha estudiat una àmplia gamma d'exposicions ambientals, inclosos el consum matern d'alcohol i de tabac, així com resultats adversos per a la salut com trastorns neurològics/cognitius i condicions cardiovasculars. En aquest context, s'ha suggerit que els fenòmens adversos ambientals contribueixen a l'ambient uterí i, en conseqüència, repercuteixen tant en el desenvolupament fetal com en la salut posterior al llarg de la vida a través de la placenta.

1.1. El paper de la placenta a DOHaD

La placenta és el primer òrgan a desenvolupar-se i té un paper clau durant l'embaràs, coordinant l'intercanvi de nutrients, gasos, residus i senyals endocrins entre la mare i el fetus. Per tant, està situat de manera única per avaluar exposicions prenatals en el context de la hipòtesi DOHaD. A més, el 2022, Bhattacharya i col·laboradors van demostrar el profund impacte en la salut de la regulació genòmica placentària en els trets de la infància, que pot persistir més tard en la vida com a antecedents etiològics per a trets complexos, per tant, programant el desenvolupament al llarg del curs de vida. En els últims anys, s'ha demostrat que les dades òmiques provinents de la placenta, inclosa l'epigenètica que és parcialment regulada per la genètica, són útils per estudiar les nombroses relacions entre l'exposició prenatal i l'estat de salut fetal i infantil.

2. L'epigenètica, el pont entre la genètica i els factors ambientals

La metilació de l'ADN (mADN) és un procés epigenètic que regula l'expressió gènica i s'ha considerat un pont entre l'entorn i el genoma, que al mateix temps, està sota el control tant de factors ambientals com genètics.

S'ha observat que la mADN pateix extenses alteracions no només durant la gestació, sinó també en resposta a les condicions de l'úter. Actualment, hi ha molts estudis que proposen que l'entorn intrauterí altera la funció placentària a través de la mADN. És important assenyalar que, mentre que la mADN es distribueix bimodalment en la majoria dels teixits i tipus cel·lulars, amb gran part dels dinucleòtids CpG hipometilats o hipermetilats, la mADN placentari segueix una distribució trimodal, principalment a causa del seu alt contingut tant en dominis parcialment metilats (PMD, per les seves sigles en anglès *partially methylated domains*) com en dinucleòtids CpG amb nivells de metilació intermedis.

No obstant això, les particularitats de la mADN placentari no només s'observen a escala general de teixit, sinó també als tipus cel·lulars. El 2021, Yuan i col·laboradors van advertir que els perfils de la mADN més particulars eren els de trofoblasts placentaris (TB), que durant la gestació es diferencien en sincitiotrofoblasts (STB). El STB és el tipus cel·lular més comú en les placentes de tercer trimestre, i forma una capa contínua, multinucleada i especialitzada de cèl·lules epitelials que orquestra la interacció biomolecular entre la mare i el fetus.

Pocs estudis o cap han investigat especialment el paper de l'epigenoma específic del tipus cel·lular en resposta a exposicions a l'embaràs i/o que condueixin a diversos efectes en la salut. L'important paper de la placenta i el seu perfil únic de la mADN, observat tant a escala general de teixit com als tipus cel·lulars, ens porta a especular que la interacció genoma-ambient i el seu impacte en la mADN són molt probablement únics de la placenta i mereixin una investigació més profunda.

2.1. Estudis d'associació de l'epigenoma complet

L'objectiu dels estudis d'associació d'epigenoma complet (EWAS, per les seves sigles en anglès *epigenome-wide association study*) és examinar la variació epigenètica a tot el genoma, predominantment en els dinucleòtids CpG, per detectar posicions (DMP, per les seves sigles en anglès *differentially methylated positions*) i regions (DMR, per les seves sigles en anglès *differentially methylated regions*) amb nivells de metilació diferencials que s'associen significativament amb fenotips d'interès.

2.1.1. Estudis del consorci internacional Pregnancy And Childhood Epigenetics

En els últims anys, el consorci *Pregnancy And Childhood Epigenetics* (PACE) del *National Institute of Health* (NIH) ha liderat diversos EWAS que demostren que la mADN de la sang perifèrica i del cordó umbilical és sensible als factors ambientals que envolten la gestació. Però, pel que fa a la mADN placentari, un nombre menor d'estudis han examinat l'efecte de les exposicions prenatals en aquest òrgan, i els que ho van fer han demostrat l'especificitat tissular dels perfils de la mADN.

A més, l'índex de massa corporal matern previ a l'embaràs (ppBMI, per les seves sigles en anglès *pre-pregnancy body mass index*) és d'especial interès, ja que un ppBMI matern més alt s'associa amb el creixement fetal aberrant, la macrosomia i l'augment de la morbiditat neonatal, així com amb complicacions de l'embaràs, inclosa la preeclàmpsia, la diabetis gestacional, la hipertensió gestacional, el part prematur i la cesària. S'ha observat que el ppBMI matern s'associa amb l'estat de salut de la descendència en la vida posterior, inclosos un major risc d'obesitat i un pitjor rendiment cognitiu en els infants.

El 2017, el consorci PACE va estudiar l'efecte del ppBMI matern en la sang del cordó umbilical, demostrant que aquesta exposició està àmpliament associada amb la mADN d'aquest teixit. No obstant això, els autors van observar que molts efectes epigenètics significatius eren modestos, i no van detectar enriquiments amb cap ruta biològica particular, deixant obertes les preguntes sobre els possibles mecanismes intrauterins que podrien estar afectant el perfil epigenètic del nounat. En aquest context, mentre que les modificacions epigenètiques en sang de cordó i sang perifèrica han estat investigades a fons, l'impacte potencial del ppBMI matern en la

mADN placentari continua sent poc explorat, i els estudis realitzats en aquest teixit van donar un nombre limitat de resultats significatius, probablement a causa del reduït nombre de mostres.

2.2. Locus de trets quantitatius de metilació

Els loci de trets quantitatius de metilació (mQTL, per les seves sigles en anglès *methylation quantitative trait locus*) són polimorfismes de nucleòtids simples (SNP, per les seves sigles en anglès *single-nucleotide polymorphism*) en els quals la variació genètica s'associa amb els nivells de la mADN d'un dinucleòtid CpG. En general, els QTLs s'han utilitzat per entendre els efectes funcionals de les variants genètiques, moltes de les quals es troben en regions no codificants, i associades a nombrosos trets i malalties. El 2021, Oliva *et al.* van caracteritzar el perfil de la mADN del genoma de diversos teixits humans sans, i van demostrar que els mQTLs poden revelar un nombre substancial d'associacions moleculars de trets que d'altra manera no es troben a les aproximacions de loci d'expressió de tret quantitatiu (eQTL, per les seves sigles en anglès *expression quantitative traït loci*), identificant possibles gens candidats.

En aquest context i com ja hem esmentat, els perfils de la mADN no sols són específics del teixit, sinó també del tipus cel·lular. D'aquesta manera, Kim-Hellmuth i els seus col·laboradors van descobrir que els QTLs que interaccionen amb el tipus cel·lular (iQTL) són altament específics de teixit i de tipus cel·lular i proporcionen una resolució més precisa que l'obtinguda amb els QTLs estàndards (sense un terme d'interacció amb un tipus cel·lular). Així, els iQTLs són instruments molt valuosos que ens permeten entendre el mecanisme de les associacions genètiques amb trets complexos.

2.2.1. Estudis del mQTL placentari

En els últims anys, s'han realitzat diversos estudis que han mapat mQTLs placentaris com Do *et al*. 2016, Delahaye *et al*. 2018, Tekola-Ayele *et al*. 2022 i Casazza *et al*. 2024. No obstant això, no hi ha una base de dades del mQTL placentari d'accés públic que mesuri la mADN en més dinucleòtids CpG que els del chip d'Illumina 450K, ni una amb un nombre de mostres considerable, o una que tingui en compte l'heterogeneïtat de tipus cel·lular i els seus perfils epigenètics específics.

3. Anàlisis d'Aleatorització Mendeliana i Colocalització

Avui dia, els estudis de l'associació de genoma complet (GWAS, per les seves sigles en anglès *genome-wide association study*) han identificat milers de variants genètiques associades a trets complexos humans. No obstant això, els gens o elements funcionals de l'ADN a través dels quals aquestes variants exerceixen els seus efectes sobre els trets romanen desconeguts. Inferir relacions causals entre fenotips és un repte important i té implicacions importants per entendre l'etiologia de la salut i la malaltia. Anàlisis estadístiques com la col·localització i l'anàlisi d'aleatorització mendeliana (MR, per les seves sigles en anglès *mendelian randomization*) poden ser útils per inferir estadísticament els efectes causals entre els dos trets, és a dir, un tret molecular (com la mADN) i una malaltia.

La col·localització analitza si dos senyals d'associació independents en el mateix locus són causades per una mateixa variant gènica causal compartida. Si la resposta és positiva, ens referim a aquesta situació com a trets que col·localitzen, de manera que la probabilitat que tots dos trets comparteixin un mecanisme causal augmenta. Un exemple comú implica un estudi d'eQTL i un resultat d'associació d'una malaltia, que apunten al gen causal i al teixit en el qual exerceix el seu efecte sobre la malaltia.

A més, la MR utilitza la variació genètica per abordar qüestions causals sobre com les exposicions modificables influeixen en els diferents resultats. Els principis de la MR es basen en les lleis d'herència de Mendel i els mètodes instrumentals d'estimació variable, que permeten la inferència d'efectes causals en presència de variables confusores no identificades. Breument, la metodologia de la MR es pot interpretar com una anàlisi per provar si l'efecte d'un SNP en un tret, com una malaltia, està mediat per una exposició, com la mADN.

4. L'associació entre els trets neuropsiquiàtrics i l'entorn prenatal

Les condicions neuropsiquiàtriques que no solen manifestar-se fins a la infància o la joventut es reconeixen cada vegada més per tenir els seus orígens en el període fetal. Com a tal, la salut materna i les exposicions ambientals intrauterines poden

influir en el desenvolupament cerebral fetal, sovint a través de vies fisiològiques convergents. D'aquesta manera, s'havien descrit com a factors de risc significatius per al desenvolupament de trastorns comuns com el trastorn de l'espectre autista (ASD, *autism spectrum disorder*), el dèficit d'atenció i el trastorn d'hiperactivitat (ADHD, *attention-deficit and hyperactivity disorder*), l'esquizofrènia (SCZ, *schizophrenia*) i el trastorn bipolar (BIP, *bipolar disorder*). Per exemple, altres condicions neuropsiquiàtriques estudiades en aquest context són trastorns conductuals com ara la internalització (INT) i externalització dels problemes, que inclouen el trastorn depressiu major (MDD, *major depressive disorder*) el trastorn obsessivocompulsiu (OCD, *obsessive-compulsive disorder*) i comportament agressiu en la infància (AGR, *aggressive behavior*).

En conclusió, cada vegada són més els estudis que demostren l'associació entre les exposicions a l'embaràs, com el ppBMI matern, i el desenvolupament posterior d'una àmplia gamma de trets i trastorns en el context de la hipòtesi DOHaD. De la mateixa manera, els estudis epigenètics han demostrat ser un pont entre els factors genètics i ambientals, tot i això la majoria d'aquests estudis s'han portat a terme en teixit sanguini perifèric i del cordó umbilical, deixant de banda el potencial paper de la placenta com un dels actors clau en un embaràs d'èxit. En els últims anys, nous estudis han demostrat que aquest òrgan té marques epigenètiques específiques i úniques, diferents de les identificades en sang de cordó umbilical o qualsevol altre teixit. Aquests patrons poden ser un dels mecanismes pels quals les exposicions prenatals i els gens de susceptibilitat exerceixen els seus efectes. En aquest sentit, s'ha suggerit que alguns trets neuropsiquiàtrics tenen orígens prenatals, però un nombre limitat d'articles s'han centrat en l'epigenoma placentari com el mecanisme potencial pel qual la genètica podria estar contribuint al desenvolupament d'aquests trastorns.

Objectius

L'objectiu general d'aquesta tesi és dilucidar l'impacte del ppBMI matern, un factor ambiental fetal rellevant, i la genètica fetal sobre la salut posterior al llarg de la vida a través de la mADN placentari.

Els objectius operatius d'aquest estudi van ser:

1. Determinar la possible associació del ppBMI matern amb la mADN placentari de tot l'epigenoma, i entendre els mecanismes pels quals l'obesitat materna podria associar-se amb el futur estat de salut de la descendència.
2. Construir i posar a disposició del públic la base de dades del mQTL placentari més gran fins al moment, com una nova eina per entendre l'efecte de la mADN placentari en l'estat de salut al llarg de la vida.
    a. Identificar i caracteritzar els dinucleòtids CpG o llocs de mADN de la placenta regulats per la genètica, i més específicament per SNPs.
    b. Identificar i caracteritzar els dinucleòtids CpG o llocs de mADN de la placenta regulats per la interacció entre SNPs i les proporcions dels tipus cel·lulars, edat gestacional (GA, *gestational age*) i sexe fetal.
3. Determinar si la mADN placentari és el mecanisme pel qual els factors de risc genètics contribueixen al desenvolupament de diverses malalties neuropsiquiàtriques, com el ADHD, l'AGR, l'ASD, el BIP, la INT, el MDD, el OCD, el trastorn de pànic (PD, *panic disorder*), l'intent suïcida (SA, *suicidal attempts*) i la SCZ.

Mètodes

1. Biòpsies de placenta i extracció d'ADN

L'estudi de Medi Ambient i Infància (INMA, per les seves sigles en castellà *Infancia y Medio Ambiente*) és un estudi de cohort de mares i fills/filles de base poblacional a Espanya que té com a objectiu estudiar el paper dels contaminants ambientals en l'aire, l'aigua i la dieta durant l'embaràs i la infància en relació amb el creixement i desenvolupament infantil. A INMA es van seguir 2,506 parelles de mares i fetus d'Astúries, Guipúscoa, Sabadell i València fins al naixement i es va recollir una selecció de 489 placentes. Es van obtenir biòpsies de la regió interna de la placenta, i es va extreure i aïllar el seu ADN.

2. Control de qualitat de les dades del genotip

El genotipatge genòmic de 509,450 variants genètiques en 397 mostres d'INMA es va realitzar amb el chip d'Illumina GSA. El control de qualitat del genotip es va dur a terme amb el programari PLINK 1.9 seguint les recomanacions estàndard. El genotip va ser imputat amb el *Michigan Imputation Server* utilitzant el panell *Haplotype Reference Consortium*, i variants amb una qualitat d'imputació r^2 per sota de 0.9, una freqüència d'al·lel menor inferior al 5%, un P valor de l'equilibri d'Hardy-Weinberg inferior a 0.5 i amb més de dos al·lels van ser eliminats. En aquesta anàlisi només es van considerar aquelles mostres amb dades de metilació. El conjunt de dades final va consistir en 368 mostres i 4,171,035 SNPs.

3. Control de qualitat de les dades de la mADN

Seguint el protocol del fabricant, es van analitzar 190 i 397 placentes amb els xips d'Illumina 450K i EPIC per l'EWAS del ppBMI matern i la base de dades de mQTL placentari, respectivament. Breument, es van normalitzar les dades de la mADN, i es van descartar mostres de baixa qualitat i sondes conflictives. En el cas del ppBMI matern, el conjunt de dades finals consistia en 168 mostres i 405,301 sondes de mADN o dinucleòtids CpG. I, en el cas dels mQTLs placentaris, el conjunt de dades finals va consistir en 368 mostres i 747,486 sondes de llocs de mADN o dinucleòtids CpG.

A més, es va estimar l'heterogeneïtat cel·lular de les mostres amb els paquets de R RefFreeCellMix i Planet per l'estudi del ppBMI matern i dels mQTLs placentaris, respectivament.

4. EWAS del ppBMI matern i la mADN placentari

Es van realitzar regressions lineals robustes per analitzar les associacions entre els valors beta de la mADN placentari a cada dinucleòtid CpG i el ppBMI matern com a variable contínua. Es van ajustar els models per a l'edat i el nivell d'educació materna, la paritat, i el tabaquisme matern durant l'embaràs. Vam executar dos models EWAS amb i sense ajust pels components cel·lulars de RefFreeCellMix.

La metaanàlisi es va realitzar amb 2,631 mostres de deu cohorts de PACE (AQUA, EARLI EDEN, Gen3G, GENEIDA, HEBC, INMA, ITU, MARBLES, NHBCS i RICHS), els protocols de control de qualitat de la mADN eren els mateixos que els descrits anteriorment per a la cohort INMA.

Finalment, es va dur a terme una anàlisis d'enriquiment funcional dels dinucleòtids CpG sensibles al ppBMI matern, així com amb la proximitat d'aquests amb SNPs associats amb trets del naixement i amb dinucleòtids CpG associats al ppBMI matern descrits en un estudi previ de sang de cordó umbilical de PACE.

5. Anàlisi del *cis*-mQTL placentari

Es van realitzar regressions lineals entre el genotip i els valors normals transformats basats en el rang de la mADN. El model es va ajustar amb el sexe del fetus, les components principals genètiques i no genètiques de la mADN (PC, per les seves sigles en anglès *principal components*), i les proporcions de tipus cel·lular. La finestra per al mapat del mQTL es va especificar com a ±0.5 Mb de cada dinucleòtid CpG analitzat.

Després, vam dur a terme diverses proves d'enriquiment amb anotacions funcionals, com ara marques de cromatina i d'histones específiques de teixits, així com conjunts de gens de les bases de dades *Gene Ontology* i *Disease Ontology*.

6. Anàlisi del *cis*-mQTL d'interacció placentari

Els imQTLs amb el tipus cel·lular amb el tipus cel·lular van ser calculats pels STB i TB de forma independent, i van consistir en models lineals que incloïen el genotip i un dels tipus cel·lulars com a terme d'interacció, ajustats pel sexe del fetus i les PCs genètiques i no genètiques de la mADN. A més, també vam calcular imQTLs de sexe i GA (sexe-imQTL i GA-imQTL, respectivament) corregint per les mateixes covariants, a més a més de les proporcions de tipus cel·lular. En el cas dels sexe-imQTLs, el sexe del fetus va ser exclòs com a covariant.

7. Anàlisis multi-òmiques per desentranyar el potencial origen placentari d'alguns trastorns neuropsiquiàtrics

Els resums estadístics dels GWAS utilitzats en aquestes anàlisis van ser els estudis d'associació més grans i disponibles públicament del ADHD, l'AGR, l'ASD, el BIP, la INT, el MDD, el OCD, el PD, el SA i la SCZ. Aquests resums estadístics van ser harmonitzats a dbSNP *build* 155 amb les dades del genotip d'INMA com a referència.

Les anàlisis de MR es van dur a terme considerant la mADN de dinucleòtids CpG com l'exposició i els trastorns neuropsiquiàtrics com el resultat, provant així l'efecte causal de la mADN placentari en aquests trastorns. A més, es va aplicar una prova d'heterogeneïtat per descartar escenaris on dos SNPs diferents en alt desequilibri de lligament estan afectant l'exposició i el resultat per vies separades. La mateixa anàlisi de MR es va aplicar a les bases de dades imQTL, però en el cas de STB- i TB-imQTLs, només es van considerar aquells que mostraven un augment de l'efecte principal del genotip amb l'abundància de tipus cel·lular.

L'anàlisi de col·localització es va dur a terme entre els loci significatius definits en el GWAS originals i els mQTLs placentaris. Aquesta anàlisi només es va realitzar amb els tres fenotips estudiats amb més detall (el BIP, el MDD i la SCZ).

Finalment, vam avaluar la influència dels dinucleòtids CpG identificats en l'expressió gènica de la placenta en una base de dades de metilació de trets quantitatius d'expressió (eQTM, per les seves sigles en anglès *expression quantitative trait methylation*) de la cohort nord-americana RICHS. L'anàlisi de l'enriquiment del conjunt de gens dels eQTMs associats a la SCZ es va realitzar utilitzant la base de dades *Reactome pathway* i la seva eina web.

8. Anàlisi condicional entre els *cis*-mQTLs placentaris i els trastorns neuropsiquiàtrics

Vam dur a terme una anàlisi condicional condicionant el principal mQTL dels dinucleòtids CpGs que superaven la MR, però no superaven la prova d'heterogeneïtat, ja que l'heterogeneïtat de vegades pot ser causada per senyals secundaris reals.

Aquesta anàlisi condicional només es portar a cap amb els tres fenotips estudiats amb més detall (el BIP, el MDD i la SCZ).

Resultats

1. El ppBMI matern està associat a 27 llocs de mADN placentari

Després d'aplicar la correcció de Bonferroni, vam obtenir 27 i 42 dinucleòtids CpG en els quals el ppBMI matern es va associar significativament amb els nivells de mADN placentari en els models ajustats i no ajustats per als components de tipus cel·lular, respectivament. L'alt ppBMI matern es va associar amb la reducció de la mADN placentari en 24/27 dinucleòtids CpG identificats en el model ajustat de tipus cel·lular, mentre que en el model no ajustat, 33/42 resultats van mostrar associacions positives. No obstant això, la mida de l'efecte dels dinucleòtids CpG que es trobaven diferencialment metilats en un model es van correlacionar positivament amb les mides de l'efecte del mateix dinucleòtid en l'altre model. Finalment, l'heterogeneïtat de les associacions entre cohorts va ser menor per al model ajustat en comparació amb el model no ajustat i, per tant, vam continuar amb els resultats del model ajustat per les proporcions cel·lulars en les anàlisis posteriors.

Entre els 27 DMPs identificats en el nostre EWAS ajustat pel tipus cel·lular, cal esmentar alguns dinucleòtids CpG de forma específica. Per exemple, el dinucleòtid CpG més significatiu, cg08219219, va mostrar que un augment de 10 unitats en el ppBMI matern s'associa amb una reducció de la mADN d'un 1.1% en aquest dinucleòtid. La mida de l'efecte positiu més gran es va observar en cg14704941, amb un increment del 2% en la mADN placentari per cada 10 unitats del ppBMI matern. Per altra banda, la mida de l'efecte negatiu més gran es va trobar en cg04724807 amb una reducció de l'1.8% en la mADN per cada 10 unitats del ppBMI matern.

Els 26 gens anotats als 27 dinucleòtids CpG sensibles al ppBMI matern van presentar un enriquiment amb rutes relacionades amb el càncer de pulmó de cèl·lules petites

i l'estrès oxidatiu. A més, aquesta mateixa llista de gens estava enriquida amb gens regulats pel factor de transcripció *ZNF217*.

També vam avaluar si els dinucleòtids CpG identificats eren propers a SNPs associats amb trets del naixement, i si el patró de la mADN associat a la placenta i el ppBMI matern era similar al que s'havia descobert en sang de cordó umbilical. D'una banda, més d'un terç dels 27 dinucleòtids CpG associats al ppBMI matern es trobaven dins d'una finestra de ±0.5 Mb de SNPs associats amb el pes en néixer. I, d'altra banda, no es va trobar cap dinucleòtid CpG associat al ppBMI matern comú entre els dos teixits.

2. Més de 200,000 dinucleòtids CpG placentaris estan regulats per variants genètiques en *cis*

Vam detectar 214,830 mQTLs placentaris basats en 214,830 i 150,649 mSites (dinculeòtids CpG) i mVariants (SNPs), respectivament. La majoria d'ells presentaven els mVariants i els mSites propers entre si, indicant que els llocs de mADN genèticament regulats són típicament propers a la variant reguladora implicada.

En general, els mSites no es trobaven ni en illes CpG i ni en promotors gènics, per altra banda, se solien trobar en regions genòmiques amb valors de metilació intermedis com les anomenades *open sea, shelf* i *shore* de les illes CpG que corresponen als marges d'aquestes. A més, hi va haver un enriquiment amb la regió de l'antigen leucocitari humà, així com regions obertes de cromatina i transcripcionalment actives específiques de placenta. Els gens propers als mSites estaven enriquits amb 219 conjunts de gens, inclosos neuropatia, trastorn de l'estat d'ànim, malaltia desmielinitzant i el BIP.

La base de dades mQTL placentari es pot trobar a l'adreça següent: https://irlab.shinyapps.io/shiny_mqtl_placenta/.

3. Detecció de mQTLs d'interacció amb els tipus cel·lulars STB i TB

En el cas dels imQTLs, vam obtenir 38 i 1 STB- i TB-imQTLs, respectivament, però cap GA- ni sexe-imQTLs significatius. L'alta quantitat de STB-imQTLs va revelar una major

potència estadística per al tipus de cèl·lula més abundant. Gran part dels STB- i TB-imQTLs eren comuns entre els dos tipus cel·lulars, i presentaven un modest impacte de la GA. Això suggereix que la transició TB-a-STB impacta en gran manera el metiloma placentari, i l'impacte modest de la GA és com a mínim parcialment independent.

4. Anàlisis multi-òmiques per desentranyar el potencial origen placentari d'alguns trastorns neuropsiquiàtrics

Amb la MR, no es van trobar resultats significatius per a molts dels trets neuropsiquiàtrics, com l'OCD o SA. Per altra banda, es van identificar dinucleòtids CpG per als trastorns el ADHD (n = 1), l'ASD (n = 1), el BIP (n = 30), el MDD (n = 28), i particularment la SCZ (n = 214). Per replicar les associacions pleotròpiques o potencialment causals identificades en les anàlisis de MR, vam dur a terme una anàlisi de col·localització a 64, 102 i 287 regions genòmiques definides en el GWAS associades amb el BIP, el MDD i la SCZ, respectivament. Pel que fa als mQTLs, de la quantitat total de 214,830 mSites amb FDR < 0.05, 12,228, 10,343 i 38,412 mSites es van considerar per el BIP, el MDD i la SCZ, respectivament. En el BIP, 7,123 parells regió-CpG, que involucraven 6,487 dinucleòtids CpG i 64 regions, donaven suport a la col·localització. En el MDD, 3,895 parells regió-CpG, que implicaven 3,850 dinucleòtids CpG i 79 regions, eren favorables a la col·localització. Finalment, en la SCZ, 21,071 parells regió-CpG, que implicaven 19,661 dinucleòtids CpG i 274 regions, van donar suport a la col·localització. Quan es va avaluar la superposició entre els resultats de la MR i els de la col·localització, dels 30 dinculeotids CpG en el BIP, 28 (93%) van mostrar evidències de col·localització amb el BIP; en el MDD, 20 dinucleòtids CpG (71%) dels 28 obtinguts amb la MR van mostrar evidència de col·localització; i dels 214 dinucleòtids CpG obtinguts en la SCZ, 198 (92%) col·localitzaven amb la SCZ. Quan es van tenir en compte els eQTMs de la cohort RICHS, dels 28 i 214 dinucleòtids CpG resultants de la MR en el MDD i la SCZ, vam trobar que 3 i 43 es correlacionaven amb els nivells d'expressió de 4 i 26 gens significatius en els eQTMs, respectivament, donant suport al rol de la placenta en els dinucleòtids identificats.

A continuació, vam identificar la intersecció entre els resultats de les tres anàlisis, la MR, la col·localització i els eQTMs. En el cas del MDD, un dinucleòtid CpG placentari

es va col·localitzar amb un locus genòmic associat el MDD al cromosoma 14, i els seus nivells de metilació es van correlacionar amb l'expressió placentària del gen *LRFN5*, expressat en cèl·lules mare de TBs que podria tenir un paper important a la placenta. En la SCZ, 40 dinucleòtids CpG placentaris van col·localitzar amb 12 loci associats a la SCZ en els cromosomes 1, 3, 6, 7, 8, 11, 12, 18, 19 i 22, i els seus nivells de metilació es van correlacionar amb l'expressió placentària de 23 gens propers, entre els 26 totals, sent aquests *GLB1L3, H2BC6, H3C4, KCNG2, LETM2, NAGA, PGBD1, PSMG3, RNF39, SDCCAG8, SFMBT1, SLC6A16, TOB2P1, TRIM27, VPS37B, ZKSCAN4, ZNF165, ZNRD1-AS1, ZSCAN12, ZSCAN12* i *ZSCAN23*.

Alguns d'aquests gens són dignes d'esment. D'una banda, l'*H2BC6* és una histona que s'ha identificat a la placenta de mares amb avortaments espontanis recurrents injustificats. D'altra banda, el *H3C4* està desregulat en embrions de ratolins durant l'exposició a l'alcohol en la neurulació primerenca a causa de canvis aberrants en els patrons de mADN. Finalment, el *VPS37B* és un gen crucial en el procés víric anomenat *budding*. En conjunt, aquests gens participen en l'homeòstasi del sistema immune, en la infecció viral i semblen ser rellevants per al desenvolupament de l'embaràs.

Finalment, amb els 26 gens de la intersecció entre els resultats de les anàlisis de MR i els eQTMs en la SCZ, vam dur a terme una anàlisi funcional amb l'eina i la base de dades *Reactome*. Vam observar un enriquiment amb 101 conjunts de gens, inclosos moltes vies de regulació epigenètica, suggerint que els al·lels que confereixen risc a patir SCZ podrien canviar els patrons epigenètics de la placenta a través de la regulació, entre d'altres, dels esmentats gens codificadors d'histones. A més, vam detectar un notable enriquiment amb vies relacionades amb la immunitat com la infecció per citomegalovirus humà i la infecció vírica general, reforçant una hipòtesi que apunta a l'activació immunitària materna (MIA, per les seves sigles en anglès *maternal immune activation*) com a vincle entre la placenta i el risc de patir SCZ.

5. Es detecta un senyal secundari potencialment causal en la SCZ

Amb l'anàlisi condicional, vam descobrir una regió en la SCZ, en la qual dos SNPs associats independentment tenien un efecte pleotròpic o potencialment causal tant en cg27314558 com en la SCZ. Aquest dinucleòtid CpG està envoltat per múltiples ARN de transferència (tARN), sent un tARN d'alanina el més proper. Se sap que la transferència placentària deficient de tARNs s'associa amb la restricció de creixement fetal.

6. Detecció d'associacions pleotròpiques entre el BIP i la SCZ, i mQTLs d'interacció amb STB i TB

La MR entre els imQTLs i el BIP, el MDD i la SCZ, va resultar en un STB-imQTL pleotropicalment associat el BIP, i dos TB-imQTLs pleotropicalment associats el BIP i el MDD (un cada un). Per exemple, el CpG implicat en el TB-imQTL pleotropicalment associat amb el BIP es troba en el cos genètic de *SMAD3*. La sobreexpressió d'aquest gen a la placenta s'ha demostrat que activa la capacitat dels TBs per formar xarxes de tipus endotelials, i s'associa amb la preeclàmpsia. A més, aquest gen és una diana ben coneguda per al tractament de liti en el BIP.

Discussió

1. L'EWAS de ppBMI matern

Aquí presentem l'EWAS més gran dut a terme en la mADN placentari fins al moment. Vam identificar 27 dinucleòtids CpG en els quals vam observar variacions en la mADN placentari del 0.5 al 2.0% per 10 unitats de diferència del ppBMI matern. A més, les nostres troballes en la mADN semblen ser específiques de la placenta, mostrant una superposició mínima amb una metaanàlisi previa de la mADN de la sang de cordó umbilical en relació amb el ppBMI matern. Els dinucleòtids CpG metilats diferencialment es troben principalment en regions de *open sea*, amb una absència en les illes CpG, i enriquits amb rutes relacionades amb el càncer i l'estrès oxidatiu.

Aquestes observacions, juntament amb el fet que el ppBMI matern s'associa amb la mADN placentari en dinucleòtids CpG situats prop de gens relacionats amb l'obesitat, ens porta a plantejar la hipòtesi que la mADN placentari podria ser un dels mecanismes pels quals l'obesitat materna s'associa amb el creixement fetal aberrant i l'estat de salut metabòlica de la descendència al llarg de la vida. No obstant això, no podem descartar que els canvis observats podrien ser només marcadors d'exposició a un alt ppBMI matern, sense efectes causals en l'estat de salut al llarg de la vida, i, per tant, les nostres troballes hauran de complementar-se amb estudis funcionals o anàlisis d'inferència causal per entendre millor si realment tenen un paper en les complicacions de l'embaràs o la salut metabòlica a llarg termini.

2. mQTLs placentaris

En conclusió, vam trobar que els mQTLs placentaris són altament específics de la placenta, amb un notable enriquiment amb regions genòmiques actives a la placenta i rutes relacionades amb el neurodesenvolupament i la salut mental. Vam demostrar que són útils per mapar l'etiologia dels trastorns neuropsiquiàtrics a estadis prenatals i vam concloure que part del risc genètic de patir el BIP, el MDD i en particular la SCZ, actuen a través de la mADN placentari en loci genòmics específics. De fet, algunes de les associacions observades poden ser causals en lloc de pleotròpiques a causa de la presència de senyals d'associació secundaris en anàlisis condicionals, la implicació d'imQTLs específics de tipus cel·lular i l'associació amb els nivells d'expressió de gens rellevants a la placenta. És d'especial interès que la mADN placentari associat a la SCZ es correlacioni amb l'expressió de gens relacionats amb la infecció viral i la regulació epigenètica, donant suport a la hipòtesi de la MIA com a peça fonamental en els orígens prenatals de la SCZ. Es necessitaran futurs estudis, inclosos enfocaments funcionals com la modulació *in vitro* de l'expressió en línies cel·lulars efectores potencials, així com un seguiment a llarg termini en estudis prospectius, per tal de caracteritzar millor les nostres troballes i la seva implicació en el desenvolupament de malalties.

Observacions finals

Les exposicions primerenques i la genètica de risc semblen tenir un impacte en la salut futura de la descendència modificant la mADN placentari. S'ha demostrat que

els patrons de la mADN placentari descrits aquí són, almenys parcialment, únics per a aquest òrgan, reforçant el seu paper en el context de la hipòtesi DOHaD. Donats els patrons específics identificats en placenta, aquest treball també destaca la importància d'aplicar en aquest òrgan els estudis prèviament realitzats en sang en el consorci PACE. Finalment, hem apuntat a diversos gens potencialment causals que, segons diferents evidències, podrien ser subjacents a la patogènesi de diferents trastorns neuropsiquiàtrics exercint els seus efectes en la placenta. Entre d'altres, la intersecció de loci pleotròpics amb eQTMs placentaris, la pleotropia regional reportada, i fins i tot la presència de senyals d'associació secundaris derivats d'anàlisis condicionals, suggereix amb fermesa que part del risc genètic de patir el BIP, el MDD i molt especialment la SCZ, actua a través de la mADN placentari en loci genòmics específics. Les perspectives futures han d'incloure estudis amb un major nombre de mostres i participants d'altres poblacions per augmentar el poder estadístic i per aturar l'exclusió sistemàtica de descendents no-Europeus.

Conclusions

Les conclusions de la tesi doctoral són:

1. El ppBMI matern impacta en la mADN placentari. En aquesta tesi hem identificat 27 dinucleòtids CpG específics de la placenta associats al ppBMI matern, molts dels quals es troben en regions *open sea*, propers a gens relacionats amb l'obesitat com *GPX1* i *LGR4* i en conjunt, enriquits amb rutes relacionades amb el càncer i l'estrès oxidatiu. Proposem que la mADN placentari podria ser un dels mecanismes pels quals l'obesitat materna s'associa amb l'estat de la salut metabòlica en els períodes postnatal i infància.
2. Hem posat a disposició de forma pública un catàleg de mQTLs placentaris, incloent-hi 214,830 dinucleòtids CpG modelats en funció del genotip dels fetus. A més, hem identificat que les proporcions de tipus cel·lular afecten la mADN placentari i hem identificat 38 i 1 dinucleòtids CpG en la placenta que són modelats en funció de les proporcions cel·lulars de STB i TB, respectivament, de manera dependent del genotip.

a. Els *cis*-mQTLs placentaris no es troben en regions normalment hipometilades, i, en canvi, es localitzen en regions amb valors intermedis de la mADN, com ara els marges de les illes CpG. A més, els *cis*-mQTLs semblen altament específics de la placenta i de manera interessant, són enriquits amb marques de cromatina activa específiques de la placenta, i amb vies relacionades amb la neuropatia i el trastorn de l'estat d'ànim, entre d'altres.

b. Observem una major potència estadística per al tipus de cèl·lula més abundant i, per tant, una major quantitat de STB-imQTLs en comparació amb TB-imQTLs. No obstant això, hi ha una superposició notable i, com s'esperava, els efectes al·lèlics en els imQTLs superposats estan negativament correlacionats entre els dos tipus de cèl·lules. Per altra banda, la moderada superposició entre els STB- i els GA-imQTLs suggereix que l'efecte escàs de la GA en la regulació genètica de la mADN placentari és almenys parcialment independent de la transició TB-a-STB.

3. Part de la predisposició genètica respecte el BIP, el MDD i particularment a la SCZ, actua a través de la mADN placentari. La causalitat potencial de diverses de les associacions observades es reforça mitjançant senyals d'associació secundaris identificats en l'anàlisi condicional, la identificació de imQTLs de tipus cel·lular, i la correlació entre els nivells de mADN dels dinucleòtids CpG identificats amb els nivells d'expressió de gens rellevants a la placenta. A més, la presència de gens relacionats amb la immunitat i l'epigenètica, inclosos *VPS37B*, *H2BC6* i *H3C4*, com a eQTMs dels dinucleòtids CpG associats pleotròpicament, reforcen els orígens prenatals de la SCZ i el paper de la MIA en aquesta. En conclusió, la predisposició genètica respecte a diversos trastorns neuropsiquiàtrics podria operar a través de la mADN placentari i l'expressió gènica associada a la placenta

RESUMEN

Introducción

1. La hipótesis de los orígenes evolutivos de la salud y la enfermedad

La hipótesis de los orígenes del desarrollo de la salud y la enfermedad (DOHaD, por sus siglas en inglés *Developmental Origins of Health and Disease*) fue acuñada por Barker en 2007 y propone que tanto el entorno perinatal como el temprano tras el nacimiento, pueden afectar la salud fetal y posterior. Se ha estudiado una amplia gama de exposiciones ambientales, incluido el alcohol y el tabaquismo materno, así como resultados adversos para la salud, como trastornos neurológicos/cognitivos y afecciones cardiovasculares. En este contexto, se ha propuesto que los estímulos ambientales contribuyen a la regulación del entorno intrauterino y, en consecuencia, afectan tanto al desarrollo fetal como a la salud posterior a través de la placenta.

1.1. El papel de la placenta en DOHaD

La placenta es el primer órgano en desarrollarse y desempeña un papel clave durante el embarazo, coordinando el intercambio de nutrientes, gases, desechos y señales endocrinas entre la madre y el feto. Por lo tanto, se encuentra en una situación única para evaluar las exposiciones prenatales en el contexto de la hipótesis DOHaD. Además, en 2022, Bhattacharya y sus colaboradores demostraron el profundo impacto causado por la regulación genómica de la placenta en rasgos complejos de la vida temprana que pueden persistir más adelante como antecedentes etiológicos de rasgos complejos de aparición más tardía y, por lo tanto, programar el desarrollo a lo largo del curso de la vida. En los últimos años, las ómicas relacionadas con la placenta, incluidos los datos epigenómicos que están regulados en parte por la genética, han demostrado ser útiles para estudiar la relación entre las exposiciones prenatales y los resultados de salud en etapa fetal e infancia temprana.

2. Epigenética, el puente entre la genética y los factores ambientales

La metilación del ADN (mADN) es un proceso epigenético que regula la expresión génica y ha sido considerado un puente entre el medio ambiente y el genoma, que a la vez está bajo el control de factores tanto ambientales como genéticos.

Se ha observado que la mADN sufre profundas alteraciones no sólo durante la gestación, sino también en respuesta a condiciones y estímulos intrauterinos. Hoy en día existen numerosos estudios que proponen que el ambiente intrauterino altera la función de la placenta a través de la mADN. Es importante señalar que, mientras que la mADN se distribuye bimodalmente en la mayoría de los tejidos y tipos celulares, con gran parte de los sitios CpG hipometilados o hipermetilados, la mADN de la placenta sigue una distribución trimodal, principalmente debido a su alto contenido en dominios parcialmente metilados (PMD, por sus siglas en inglés *partially methylated domains*) y posiciones CpG con niveles de mADN intermedios.

Sin embargo, las particularidades de la mADN de la placenta no sólo se observan en las biopsias de tejido, sino también en los tipos celulares que las componen. En 2021, Yuan y sus colaboradores observaron que los perfiles de mADN más característicos eran los de los trofoblastos (TB), que durante la gestación se diferencian en sincitiotrofoblastos (STB). Los STB son el tipo de célula más común en las placentas a término y forman una capa continua, multinucleada y especializada de células epiteliales que orquesta la interacción biomolecular entre la madre y el feto.

Hasta la fecha, ningún estudio ha investigado el papel del epigenoma específico de tipo celular en respuesta a exposiciones durante el embarazo y/o como mecanismo causal en diversos resultados de salud. El importante papel de la placenta y el perfil único de la mADN observado tanto a nivel de tejido como de tipo celular en este órgano, nos llevan a especular que la interacción genoma-ambiente y su impacto en la mADN es muy probablemente único en la placenta y, por lo tanto, merece ser investigado en profundidad.

2.1. Estudios de asociación de epigenoma completo

El objetivo de los estudios de asociación de epigenoma completo (EWAS, por sus siglas en inglés *epigenome-wide association study*) es examinar la variación epigenética de todo el genoma, predominantemente la mADN en sitios CpG, para detectar posiciones (DMP, por sus siglas en inglés *differentially methylated positions*) y regiones (DMR, por sus siglas en inglés *differentially methylated regions*) diferencialmente metiladas que estén significativamente asociadas con fenotipos de interés.

2.1.1. Estudios del consorcio internacional Pregnancy And Childhood Epigenetics

En los últimos años, el consorcio *Pregnancy And Childhood Epigenetics* (PACE) del *National Institute of Health* (NIH) de Estados Unidos ha liderado varios EWAS que demuestran que tanto la mADN de la sangre periférica como la de la sangre del cordón umbilical es sensible a los factores ambientales que rodean la gestación. Sin embargo, con respecto a la mADN de la placenta, un número menor de estudios ha examinado el efecto de las exposiciones prenatales en este órgano, y los que lo hicieron han demostrado la especificidad tisular de los patrones de la mADN.

Por otro lado, el índice de masa corporal al inicio del embarazo (ppBMI, por sus siglas en inglés *pre-pregnancy body mass index*) materno es de particular interés ya que un ppBMI materno más alto se asocia con un crecimiento fetal aberrante, macrosomía y una mayor morbilidad neonatal, así como con complicaciones del embarazo, incluida la preeclampsia, la diabetes gestacional, la hipertensión gestacional, el parto prematuro y la cesárea. Además, se ha observado que el ppBMI materno se asocia con otros resultados de salud de los hijos e hijas en etapas del desarrollo más avanzadas, incluidos un mayor riesgo de obesidad y un peor rendimiento cognitivo.

En 2017, el consorcio PACE estudió el efecto del ppBMI materno en la sangre del cordón umbilical y demostró que esta exposición está ampliamente asociada con la mADN en este tejido en particular. Sin embargo, los autores observaron que muchos efectos epigenéticos significativos eran modestos y no detectaron enriquecimientos para ninguna vía biológica en particular, dejando preguntas abiertas sobre los posibles mecanismos intrauterinos que podrían estar detrás de los perfiles epigenéticos identificados. Así, si bien las alteraciones epigenéticas en la sangre del cordón

umbilical y la sangre periférica se han investigado a fondo, el impacto potencial del ppBMI materno en la mADN de la placenta sigue siendo un terreno poco explorado, y los estudios realizados en este tejido arrojaron un número limitado de resultados significativos, probablemente debido a sus tamaños de muestra limitados.

2.2. Locus de rasgos cuantitativos de metilación

Los loci de rasgos cuantitativos de metilación (mQTL, por sus siglas en inglés m*ethylation quantitative trait locus*) son polimorfismos de un solo nucleótido (SNP, por sus siglas en inglés *single-nucleotide polymorphism*) en los que la variación genética está asociada con los niveles de mADN de un dinucleótido CpG. En general, los QTLs se han utilizado para comprender los efectos funcionales de variantes genéticas, muchas de las cuales están ubicadas en regiones no codificantes y asociadas con numerosos rasgos y enfermedades. En 2021, Oliva *et al.* caracterizaron el perfil de la mADN de todo el genoma en diversos tejidos humanos sanos y demostraron que los mQTLs pueden revelar una cantidad sustancial de relaciones moleculares con rasgos y enfermedades que, de otro modo, pasarían inadvertidas. Por tanto, se identificaron genes candidato potencialmente causales que no se habrían podido hallar utilizando eQTLs.

En este contexto y como ya hemos mencionado, los perfiles de mADN no sólo son específicos de tejido, sino también de tipo celular. En este sentido, Kim-Hellmuth y sus colaboradores descubrieron que los QTLs que interactúan con el tipo de célula (iQTL) están fuertemente enriquecidos en cuanto a especificidad tisular y celular, y proporcionan una resolución más precisa de la especificidad tisular que la de los QTL estándares (sin un término de interacción con un tipo celular). Por lo tanto, los iQTLs son instrumentos muy valiosos que nos permiten obtener una comprensión mecanicista de las asociaciones genéticas con rasgos complejos.

2.2.1. Estudios de mQTLs en placenta

En los últimos años, se han realizado varios estudios que han mapeado los mQTLs de placenta, como los de Do *et al.* 2016, Delahaye *et al.* 2018, Tekola-Ayele *et al.* 2022 y Casazza *et al.* 2024. Sin embargo, no existe una base de datos de mQTLs de placenta de acceso público con información sobre los niveles de mADN en CpGs más allá del chip

de metilación Illumina 450K, ni una con un tamaño de muestra considerable, o una que tenga en cuenta la heterogeneidad de tipo celular y sus perfiles epigenéticos específicos.

3. Análisis de Aleatorización Mendeliana y Colocalización

Hoy en día, los estudios de asociación de todo el genoma (GWAS, por sus siglas en inglés *genome-wide association studies*) han identificado miles de variantes genéticas asociadas con rasgos complejos humanos. Sin embargo, aún se desconocen los genes o elementos funcionales del ADN a través de los cuales estas variantes ejercen sus efectos sobre dichos rasgos. Inferir relaciones causales entre fenotipos es un gran desafío y tiene implicaciones importantes para comprender la etiología de las enfermedades. Enfoques como los análisis de aleatorización mendeliana (MR, por sus siglas en inglés *mendelian randomization*) y de colocalización pueden ser útiles para inferir estadísticamente los efectos causales entre dos rasgos, es decir, entre un rasgo molecular (mADN) y una enfermedad.

La MR utiliza la variación genética para abordar cómo las exposiciones modificables influyen en los diferentes resultados. La MR se basa en las leyes de herencia de Mendel y en los métodos de estimación de variables instrumentales, que permiten inferir efectos causales en presencia de factores de confusión no observados. Brevemente, la metodología de la MR puede interpretarse como un análisis para probar si el efecto de un SNP en un resultado, como una enfermedad, está mediado por una exposición, como la mADN.

Por otro lado, los análisis de colocalización prueban si dos señales de asociación independientes en el mismo locus son consistentes y podrían ser explicadas por una única variante causal compartida. Si la respuesta es positiva, nos referimos a esta situación como rasgos colocalizados, y la probabilidad de que ambos rasgos compartan un mecanismo causal aumenta enormemente.

4. La asociación entre los rasgos neuropsiquiátricos y el entorno prenatal

Cada vez hay más evidencias de que las afecciones neuropsiquiátricas que normalmente no se manifiestan hasta la niñez o la edad adulta temprana tienen su origen en el período fetal. Como tal, la salud materna y las exposiciones ambientales

intrauterinas pueden influir en el desarrollo del cerebro fetal, a menudo a través de vías fisiológicas convergentes. Por lo tanto, se han descrito estos factores como factores de riesgo importantes para el desarrollo de trastornos comunes como el trastorno del espectro autista (ASD, *autism spectrum disorder*), el trastorno por déficit de atención e hiperactividad (ADHD, *attention-deficit hyperactivity disorder*), la esquizofrenia (SCZ, *schizophrenia*) y el trastorno bipolar (BIP, *bipolar disorder*). Por otro lado, otras condiciones neuropsiquiátricas estudiadas en este contexto son trastornos conductuales como la internalización (INT, *internalizing problems*) y externalización de problemas, que incluyen el trastorno de depresión mayor (MDD, *major depression disorder*), el trastorno obsesivo-compulsivo (OCD, *obsesive compulsive disorder*) y el comportamiento agresivo en la infancia (AGR, *aggressive behavior*).

En conclusión, cada vez más estudios muestran la asociación entre las exposiciones durante el embarazo, como el ppBMI materno, y el desarrollo posterior de una amplia gama de rasgos y trastornos en el contexto de la hipótesis DOHaD. De manera similar, la epigenética parece funcionar como puente entre la genética y los factores ambientales durante el embarazo, pero la mayoría de las evidencias vienen de estudios realizados en sangre periférica y del cordón umbilical, dejando de lado el papel potencial de la placenta como uno de los actores clave en un embarazo exitoso. En los últimos años, novedosos estudios han demostrado que este órgano tiene marcas epigenéticas específicas y únicas, diferentes a las identificadas en la sangre del cordón umbilical o en cualquier otro tejido. Estos patrones podrían ser uno de los mecanismos por los cuales las exposiciones prenatales y los genes de susceptibilidad ejercen sus efectos. En este sentido, se ha sugerido que algunos rasgos neuropsiquiátricos tienen orígenes prenatales, pero sólo unos pocos trabajos se han centrado en el epigenoma de la placenta como el mecanismo potencial por el cual la genética podría estar contribuyendo al desarrollo de estos trastornos.

Objetivos

El objetivo general de esta tesis es dilucidar el impacto del ppBMI materno, un factor clave en el desarrollo del feto, y la genética fetal en la salud futura a través de la mADN de la placenta.

Los objetivos operativos del presente estudio fueron:

1. Determinar la posible asociación del ppBMI materno con la mADN de la placenta, y comprender los mecanismos por los cuales la obesidad materna podría estar asociada con el futuro estado de salud en la descendencia.
2. Construir y poner a disposición del público la mayor base de datos del mQTL de placenta realizada hasta la fecha, como una herramienta novedosa para entender el efecto de la mADN de la placenta en el estado de salud a lo largo de la vida.
    a. Identificar y caracterizar los sitios de mADN de la placenta regulados por la genética adyacente, y más específicamente por SNPs.
    b. Identificar y caracterizar los sitios de mADN de la placenta regulados por la interacción entre SNPs y proporciones de tipo celular, edad gestacional (GA, *gestational age*) y sexo fetal.
3. Determinar si la mADN de la placenta es el mecanismo por el cual los factores de riesgo genéticos contribuyen al desarrollo de varias enfermedades neuropsiquiátricas, incluyendo ADHD, AGR, ASD, BIP, INT, MDD OCD, trastorno de pánico (PD, *panic disorder*), intento de suicidio (SA, *suicidal attempt*) y SCZ.

Métodos

1. Biopsias de placenta y extracción del ADN

El estudio Infancia y Medio Ambiente (INMA) es un estudio de cohorte de madres e hijos e hijas en España, que tiene el objetivo de estudiar el papel de los contaminantes ambientales del aire, el agua y la dieta durante el embarazo y la infancia, en relación con el crecimiento y desarrollo infantil. En INMA, se siguió a 2,506 parejas de madres y fetos de Asturias, Gipuzkoa, Sabadell y Valencia hasta el nacimiento, y se recogieron 489 placentas. Se obtuvieron biopsias de la región interna de la placenta, y su ADN fue extraído y aislado.

2. Control de calidad del genotipo

El genotipado a nivel de genoma completo de 509,450 variantes genéticas en 397 muestras de INMA se realizó con el chip de Illumina GSA. El control de calidad del genotipo se llevó a cabo con el programa PLINK 1.9 siguiendo las recomendaciones estándar. El genotipo se imputó con el *Michigan Imputation Server* utilizando el panel *Haplotype Reference Consortium*, y las variantes con una calidad de imputación r^2 inferior a 0.9, una frecuencia del alelo menor inferior al 5%, un P valor en el equilibrio de Hardy-Weinberg inferior a 0.5, y con más de dos alelos se eliminaron. Solamente aquellas muestras que disponían de datos de mADN fueron consideradas en este análisis. El conjunto final de datos consistió en 368 muestras y 4,171,035 SNPs.

3. Control de calidad de los datos de mADN

Siguiendo el protocolo del fabricante, 190 y 397 placentas fueron analizadas con los chips de metilación de Illumina 450K y EPIC para el EWAS del ppBMI materno y la base de datos del mQTL de la placenta, respectivamente. Brevemente, se normalizaron los datos de la mADN y se descartaron muestras de baja calidad y sondas conflictivas. En el caso del ppBMI materno, el conjunto de datos final consistió en 168 muestras y 405,301 sondas o sitios CpG. Y, en el caso de los mQTLs de la placenta, el conjunto de datos final consistió en 368 muestras y 747,486 sondas o sitios CpG.

Además, la heterogeneidad celular de las muestras se estimó con los paquetes de R RefFreeCellMix y Planet para los estudios de ppBMI materno y mQTLs de la placenta, respectivamente.

4. EWAS del ppBMI materno y la mADN de la placenta

Se realizaron regresiones lineales robustas para testar las asociaciones entre los valores beta de la mADN de la placenta en cada sitio CpG y el ppBMI materno como variable continua. Los modelos fueron ajustados por edad y nivel de estudios maternos, paridad y tabaquismo materno durante el embarazo. Se ejecutaron dos modelos de EWAS con y sin ajuste por componentes celulares estimados mediante RefFreeCellMix.

El metaanálisis se realizó con 2,631 muestras de diez cohortes de PACE (AQUA, EARLI, EDEN, Gen3G, GENEIDA, HEBC, INMA, ITU, MARBLES, NHBCS y RICHS), cuyos protocolos de control de calidad de la mADN fueron idénticos a los descritos para la cohorte INMA.

Finalmente, se estimó el enriquecimiento funcional de los sitios CpG sensibles al ppBMI materno. También se estudió su proximidad con los SNPs asociados con rasgos del nacimiento y con los sitios CpG asociados al ppBMI materno reportados en un estudio previo de PACE en sangre de cordón umbilical.

5. Análisis de *cis*-mQTLs de placenta

Se realizaron regresiones lineales entre el genotipo y los valores normales de mADN. El modelo se ajustó con el sexo del feto, los componentes principales (PC, por sus siglas en ingles *principal components*) genéticos y no genéticos y las proporciones de tipo celular. La ventana para el mapeo de mQTLs fue de ±0.5 Mb a partir de cada posición CpG.

Posteriormente, realizamos varias pruebas de enriquecimiento con anotaciones funcionales, tales como la cromatina específica de tejido y diferentes marcas de histona, así como conjuntos de genes de las bases de datos *Gene Ontology* y *Disease Ontology*.

6. Análisis de *cis*-mQTLs de interacción

Los mQTLs que interaccionan con el tipo celular (imQTL) se calcularon para los tipos celulares STB y TB, y consistieron en modelos lineales que incluían el genotipo y un tipo celular como término de interacción cada vez, ajustado por el sexo del feto y las PCs genéticas y no genéticas de mADN. Además de eso, también se calcularon imQTLs de sexo fetal y de GA (GA-imQTL), corrigiendo por las mismas covariables más las estimaciones del tipo celular. En el caso de las imQTLs de sexo, el sexo fetal fue excluido de la lista de covariables.

7. Enfoques multiómicos para desentrañar el posible origen en la placenta de algunos trastornos neuropsiquiátricos

Los estadísticos de resumen de los GWAS utilizados en este trabajo fueron los mayores estudios de asociación disponibles públicamente de ADHD, AGR, ASD, BIP, INT, MDD, OCD, PD, SA y SCZ. Estos estadísticos fueron armonizados con respecto al *build* 155 de dbSNP con los datos de genotipo de INMA como referencia.

Los análisis de MR se realizaron considerando la mADN de cada CpG como la exposición y los trastornos neuropsiquiátricos como resultado, comprobando así el efecto causal de la mADN de la placenta sobre estos rasgos. Además, se aplicó una prueba de heterogeneidad en la que se consideró la posibilidad de que dos SNPs diferentes en fuerte desequilibrio de ligamiento estuvieran afectando la exposición y el resultado por separado. El mismo análisis de MR se aplicó en las bases de datos de imQTLs, pero en el caso de imQTLs de TB y STB, solo se consideraron aquellos que mostraron un aumento del efecto genotípico principal en función de la abundancia de cada tipo celular considerado.

El análisis de colocalización se realizó entre los loci significativos definidos en el GWAS original y los mQTLs de la placenta. Este análisis solo se realizó con los tres fenotipos estudiados en mayor detalle (BIP, MDD y SCZ).

Finalmente, evaluamos la influencia de los sitios CpG identificados en la expresión génica en la placenta, en una base de datos de metilación de rasgos cuantitativos de expresión (eQTM, por sus siglas en inglés *expression quantitative trait methylation*) de la cohorte norteamericana RICHS. El análisis de enriquecimiento de conjuntos de genes de los eQTMs asociados a la SCZ se realizó utilizando la base de datos de *Reactome* y su herramienta web.

8. Análisis condicional entre *cis*-mQTLs de la placenta y trastornos neuropsiquiátricos

Se realizó un análisis condicional condicionando por los mQTLs más significativamente asociados en aquellas sondas que superaron la MR pero que no pasaron la prueba de

heterogeneidad, debido a que la heterogeneidad a veces puede derivar de señales secundarias reales. Este análisis condicional solo se realizó con los tres fenotipos estudiados en mayor detalle (BIP, MDD y SCZ).

Resultados

1. El ppBMI materno se asocia con 27 sitios CpG en la placenta

Después de aplicar la corrección de Bonferroni, obtuvimos 27 y 42 sitios CpG en los que el ppBMI materno se asoció significativamente con los niveles de mADN de la placenta en los modelos ajustado y no ajustado por los componentes de tipo celular, respectivamente. Un elevado ppBMI materno se asoció con un nivel menor de mADN de la placenta en 24/27 de los sitios CpG diferencialmente metilados identificados en el modelo ajustado por tipo celular, mientras que en el modelo no ajustado, 33/42 de los CpGs mostraron asociaciones positivas. Sin embargo, el tamaño de efecto de los sitios CpG diferencialmente metilados en un modelo correlacionaban positivamente con los tamaños de efecto de la misma posición en el otro modelo. Finalmente, la heterogeneidad de las asociaciones entre las cohortes fue menor para el modelo ajustado por las proporciones de tipo celular en comparación con el modelo no ajustado y, por lo tanto, continuamos sólo con los resultados del modelo ajustado para realizar el resto de los análisis.

Entre las 27 DMPs identificadas en el EWAS ajustado por tipo celular, algunos sitios CpG son dignos de mención. Por ejemplo, el resultado más significativo, cg08219219, mostró que un aumento de 10 unidades de ppBMI materno se asocia con una mADN 1.1% menor en ese sitio CpG específicamente. El mayor tamaño positivo del efecto se observó en la posición cg14704941, con un 2% más de mADN de la placenta por 10 unidades de ppBMI materno. El mayor tamaño negativo del efecto se encontró en cg04724807, con una mADN 1.8% menor por cada 10 unidades de ppBMI.

Los 26 genes anotados a los 27 sitios CpG sensibles a ppBMI materno presentaron un enriquecimiento en la ruta del cáncer de pulmón de células pequeñas y en la vía de señalización inducida por el estrés oxidativo. Además, esa misma lista estaba enriquecida en genes regulados por el factor de transcripción *ZNF217*.

También se evaluó si los sitios CpG sensibles a ppBMI materno estaban cerca de los SNPs asociados a rasgos del nacimiento, y si el perfil de la mADN del ppBMI materno en placenta era consistente con asociaciones en sangre del cordón umbilical previamente reportadas por el consorcio PACE. Por un lado, más de un tercio de los 27 sitios CpG asociados al ppBMI estaban dentro de ±0.5 Mb de los SNPs asociados con peso al nacer. Y, por otro, no hallamos sitios CpG comunes en los dos tejidos.

2. Más de 200.000 sitios CpG están regulados por variantes genéticas en *cis* en la placenta

Se detectaron 214,830 mQTLs de la placenta basados en 214,830 mSites (sitios CpG) y 150,649 mVariants (SNPs). En la mayoría de los casos las mVariants y los mSites estaban próximos unos de otros, lo que indica que la mADN genéticamente modulada se sitúa típicamente cerca de la variante reguladora implicada.

En general, había menos mSites de los esperados en islas CpG y promotores, y más en cuerpos de genes y regiones genómicas con valores intermedios de metilación como las llamadas *open sea, shelf* y *shore* de las islas CpGs que corresponden a los márgenes de estas. Además, hubo un enriquecimiento significativo en la región del antígeno leucocitario humano, así como marcas de cromatina abierta y regiones transcripcionalmente activas específicas de placenta. Los genes cercanos a los mSites resultaron enriquecidos en 219 conjuntos de genes, incluyendo la neuropatía, trastorno del estado de ánimo, enfermedad desmielinizante y BIP.

La base de datos de mQTLs de placenta puede encontrarse en la siguiente dirección: https://irlab.shinyapps.io/shiny_mqtl_placenta/.

3. Detección de mQTLs de interacción con proporciones de tipo celular para STB y TB

En el caso de los imQTLs, obtuvimos 38 y 1 STB- y TB-imQTLs, respectivamente, sin imQTLs significativos para GA y sexo fetal. La mayor cantidad de STB-imQTLs reveló mayor poder estadístico para el tipo celular más abundante. La coincidencia entre STB- y TB-imQTLs fue notable, con un impacto modesto de la GA. Esto sugiere que la transición TB-a-STB afecta en gran medida el metiloma de la placenta, y el modesto impacto de la GA es al menos parcialmente independiente.

4. Los análisis multiómicos desentrañan el posible origen en la placenta de algunos trastornos neuropsiquiátricos

Con la MR, no se obtuvieron resultados significativos para muchos de los rasgos, como OCD o SA. Sin embargo, se identificaron sitios CpG para ADHD (n = 1), ASD (n = 1), BIP (n = 30), MDD (n = 28) y particularmente SCZ (n = 214). Para replicar las asociaciones pleiotrópicas identificadas en los análisis de MR, se realizó el análisis de colocalización en 64, 102 y 287 regiones genómicas definidas en los GWAS de BIP, MDD y SCZ, respectivamente. En cuanto a los mQTLs, del total de 214,830 mSites con FDR < 0.05, 12,228, 10,343 y 38,412 fueron considerados para BIP, MDD y SCZ, respectivamente. En BIP, 7,123 pares región-CpG, involucrando 6,487 sitios CpG y 64 regiones, obtuvieron evidencias de colocalización. En MDD, 3,895 pares región-CpG, que involucraron 3,850 sitios CpG y 79 regiones, obtuvieron evidencias de colocalización. Finalmente, en SCZ, 21,071 pares región-CpG, involucrando 19,661 sitios CpG y 274 regiones, obtuvieron evidencias de colocalización. Cuando se evaluó el solapamiento con los resultados de MR, de los 30 resultados de SMR en BIP, 28 CpGs (93%) mostraron evidencias de colocalización con BIP; en MDD, 20 sitios CpG (71%) de los 28 resultados de SMR mostraron evidencias de colocalización; y de los 214 resultados de SMR en SCZ, 198 sitios CpG (92%) mostraron evidencias de colocalización con SCZ.

Cuando se consideraron los eQTMs de la cohorte RICHS, de los 28 y 214 sitios CpG significativos en SMR para MDD y SCZ, se encontró que 3 y 43 se correlacionaban

significativamente con los niveles de expresión de 4 y 26 genes, respectivamente, apoyando la función de los sitios CpG identificados en la placenta.

A continuación, se identificó la intersección entre MR, colocalización y eQTMs para cada trastorno. En el caso del MDD, descubrimos un sitio CpG situado en un locus genómico asociado a MDD en el cromosoma 14, con sus niveles de mADN asociados con la expresión de *LRFN5* en la placenta. Éste es un gen expresado en células madre de TBs que podría desempeñar un papel importante en la placenta. En SCZ, identificamos 40 sitios CpG situados en 12 loci asociados a SCZ en los cromosomas 1, 3, 6, 7, 8, 11, 12, 18, 19 y 22, con sus niveles de mADN asociados con la expresión en placenta de 23 genes cercanos, de entre los 26 genes totales resultantes del análisis de eQTMs, concretamente *GLB1L3, H2BC6, H3C4, KCNG2, LETM2, NAGA, PGBD1, PSMG3, RNF39, SDCCAG8, SFMBT1, SLC6A16, TOB2P1, TRIM27, TXNL4A, VARS2, VPS37B, ZKSCAN4, ZNF165, ZNRD1-AS1, ZSCAN12, ZSCAN12P1,* y *ZSCAN23*.

Algunos de esos genes merecen ser mencionados. Por un lado, *H2BC6* es una histona que ha sido identificada en las placentas de madres con abortos espontáneos recurrentes injustificados. Por otro lado, el gen *H3C4* está desregulado en embriones de ratones con exposición al alcohol en la neurulación temprana, debido a cambios aberrantes en los patrones de mADN. Finalmente, *VPS37B* es un gen crucial en el proceso vírico llamado *budding*. En conjunto, esos genes participan en la homeostasis inmune, en la infección viral y parecen ser relevantes para el desarrollo del embarazo.

Finalmente, realizamos un análisis de conjunto de genes *Reactome* en los 26 genes de la intersección entre los resultados de SMR y del análisis de eQTMs en SCZ. Observamos el enriquecimiento de 101 conjuntos de genes, incluyendo muchas vías reguladoras epigenéticas, lo que sugiere que los alelos de riesgo de padecer SCZ podrían cambiar la base epigenética de la placenta a través de la regulación de, entre otros, los genes codificantes de histonas ya mencionados. Además, observamos un notable enriquecimiento en vías relacionadas con el sistema inmune, como la infección por citomegalovirus humano y la infección vírica general, reforzando la hipótesis que apunta a la activación inmune materna (MIA, por sus siglas en inglés *maternal immune activation*) como un vínculo entre la placenta y el riesgo de SCZ.

5. Detección de una señal secundaria potencialmente causal en SCZ

Con el análisis condicional, descubrimos una región en la SCZ, en la que dos SNPs asociados de forma independiente mantenían un efecto pleiotrópico tanto sobre la CpG cg27314558 como sobre la SCZ. Este sitio CpG está rodeado por múltiples ARNs de transferencia (tARNs), siendo un tARN de alanina el más cercano. Se sabe que la transferencia deficitaria de tARNs en la placenta está asociada a la restricción del crecimiento fetal.

6. Detección de asociaciones pleiotrópicas entre BIP y SCZ, y STB- y TB-imQTLs

La prueba MR entre imQTLs de tipo celular y BIP, MDD y SCZ, proporcionó un STB-imQTL pleiotrópicamente asociado a BIP, y dos TB-imQTLs pleiotrópicamente asociados a BIP y MDD (uno a cada enfermedad). Por ejemplo, la CpG involucrada en el TB-imQTL asociada a BIP está localizada en el cuerpo génico de *SMAD3*. La sobreexpresión de este gen en la placenta ha demostrado activar la capacidad de los TBs para formar redes de tipo endoteliales, y está asociada a la preeclampsia. Además, este gen es una diana bien conocida para el tratamiento con litio de BIP.

Discusión

1. EWAS de ppBMI materno

Aquí presentamos el mayor EWAS realizado en la mADN de la placenta llevado a cabo hasta la fecha. Identificamos 27 sitios CpG en los que observamos variaciones de la mADN del 0.5 al 2.0% por cada 10 unidades de ppBMI materno. Además, nuestros hallazgos parecen ser específicos de placenta, mostrando una superposición mínima con un meta-análisis anterior en mADN de sangre de cordón. Los sitios CpG diferencialmente metilados están ubicados principalmente en regiones genómicas de *open sea*, ausentes por completo en las islas CpG, y enriquecidos en rutas relacionadas con el cáncer y el estrés oxidativo.

Estas observaciones, junto con el hecho de que el ppBMI materno se asocia con la mADN de la placenta en sitios CpG ubicados cerca de genes relacionados con la obesidad, nos lleva a la hipótesis de que la mADN de la placenta podría ser uno de los mecanismos por los cuales la obesidad materna se asocia con un crecimiento fetal aberrante y tal vez, con el estado de la salud metabólica de la descendencia a lo largo de la vida. Sin embargo, no podemos descartar que los cambios observados puedan ser sólo marcadores de exposición a un elevado ppBMI materno, sin efectos causales en el estado de salud a lo largo de la vida, y, por lo tanto, nuestros hallazgos necesitarán ser complementados con estudios funcionales o análisis de inferencia causal para comprender mejor si realmente tienen un papel en las complicaciones del embarazo o en los resultados metabólicos a largo plazo.

2. mQTLs de placenta

En conclusión, encontramos que los mQTLs de placenta son altamente específicos de este tejido, con un notable enriquecimiento en regiones genómicas activas en la placenta, y vías relacionadas con el neurodesarrollo y la salud mental. Demostramos que son útiles para mapear la ventana etiológica de los trastornos neuropsiquiátricos a las etapas prenatales y concluimos que parte del riesgo genético de padecer BIP, MDD y en particular SCZ, actúan a través de la mADN de la placenta en loci genómicos específicos. De hecho, algunas de las asociaciones observadas pueden ser causales y no pleiotrópicas, debido a la presencia de señales secundarias de asociación en análisis condicionales, la implicación de QTLs específicos de tipo celular y la asociación con los niveles de expresión de genes relevantes en la placenta. Es de especial interés que la mADN de la placenta asociado a la SCZ se correlacione con la expresión de genes involucrados con la infección viral y la regulación epigenética, proporcionando más apoyo a la hipótesis de la MIA como motor en los orígenes prenatales de la SCZ. Futuros trabajos, incluyendo enfoques funcionales como la modulación *in vitro* de la expresión en líneas celulares efectoras potenciales, así como el seguimiento a largo plazo en estudios prospectivos, serán necesarios para caracterizar mejor nuestros hallazgos y su implicación en el desarrollo de la enfermedad.

Observaciones finales

Las exposiciones tempranas y la genética de riesgo parecen tener un impacto en la salud futura de los niños y niñas, a través de la mADN de la placenta. Los patrones de metilación aquí descritos han demostrado ser, al menos parcialmente, únicos para ese órgano, reforzando su papel en el contexto de la hipótesis DOHaD. Este trabajo también destaca la importancia de realizar estudios previos del consorcio PACE en la placenta, dada la especificidad de los patrones de la mADN identificados. Por último, hemos señalado varios genes potencialmente causales que, según diferentes evidencias, podrían subyacer a la patogénesis de diferentes trastornos neuropsiquiátricos al ejercer sus efectos en la placenta. Entre otros, la intersección de loci pleiotrópicos con eQTMs de placenta, la pleiotropía regional reportada, e incluso la presencia de señales secundarias de asociación derivadas de análisis condicionales, sugiere fuertemente que parte del riesgo genético de padecer BIP, MDD y muy especialmente SCZ, actúa a través de la mADN de la placenta en loci genómicos específicos. Las perspectivas futuras necesitan incluir estudios con tamaños muestrales más grandes y participantes de otras poblaciones, para aumentar el poder estadístico y detener la exclusión sistemática de población no-Europea.

Conclusiones

Las conclusiones de esta tesis doctoral son:

1. El ppBMI materno afecta a la mADN de la placenta. En esta tesis hemos identificado 27 sitios de mADN específicos de placenta asociados con el ppBMI materno, muchos de los cuales están ubicados en regiones de mar abierto, cerca de genes relacionados con la obesidad como *GPX1* y *LGR4,* y en conjunto, enriquecidos en cáncer y vías oxidativas de estrés. Proponemos que la mADN de la placenta podría ser uno de los mecanismos por los cuales la obesidad materna se asocia con resultados metabólicos en la descendencia.
2. Hemos hecho público un catálogo de mQTLs de placenta, incluyendo 214,830 sitios únicos de mADN asociados al genotipo de los fetos. Además, hemos

encontrado que las proporciones de tipo celular afectan a la mADN de la placenta, concretamente en 38 y 1 CpGs asociados no sólo al genotipo sino también a la abundancia de los principales tipos celulares, STB y TB, respectivamente.

   a. Los *cis*-mQTLs de la placenta están ausentes de regiones generalmente hipometiladas, y a su vez, enriquecidas en regiones con valores intermedios de mADN, como las regiones flanqueantes de las islas CpG. Además, los *cis*-mQTLs parecen altamente específicos de placenta y están enriquecidos en marcas de cromatina activa específicas de placenta, así como en vías relacionadas con la neuropatía y la salud mental.

   b. Se observa un mayor poder estadístico para el tipo celular más abundante y, por lo tanto, una mayor cantidad de STB-imQTLs en comparación con TB-imQTLs. Sin embargo, hay un notable solapamiento y, como se esperaba, los efectos alélicos en los imQTLs superpuestos se correlacionan negativamente entre los dos tipos celulares. El solapamiento moderado entre STB- y GA-imQTLs sugiere que el escaso efecto de la GA en la regulación genética de la mADN es al menos parcialmente independiente de la transición TB-a-STB.

3. Parte de la susceptibilidad genética a sufrir BIP, MDD y, en particular SCZ, actúa a través de la mADN de la placenta. La causalidad potencial de varias de las asociaciones observadas se ve reforzada por señales secundarias de asociación identificadas en el análisis condicional, la implicación de los tipos celulares y la correlación de las CpGs identificadas con los niveles de expresión de genes relevantes en la placenta. Además, la presencia de genes relacionados con procesos inmunitarios y epigenéticos, incluyendo *VPS37B*, *H2BC6* y *H3C4*, refuerzan los orígenes prenatales de la SCZ y el papel de la MIA. En conclusión, el riesgo genético a padecer varios trastornos neuropsiquiátricos podría operar a través de la mADN y la expresión génica asociada en la placenta.

SUMMARY

Introduction

1. The developmental origins of health and disease hypothesis

The Developmental Origins of Health and Disease (DOHaD) hypothesis was coined by Barker in 2007, and it proposes that perinatal and early life environment can impact fetal and later life health. A wide range of environmental exposures have been studied, including maternal alcohol and smoking, as well as adverse health outcomes such as neurological/cognitive disorders and cardiovascular conditions. In this context, it has been proposed that environmental insults contribute to the uterus environment and consequently impact both fetal development and later life health through placenta.

1.1. The role of placenta in DOHaD

Placenta is the first organ to develop and plays a key role throughout pregnancy, coordinating the exchange of nutrients, gases, waste and endocrine signals between mother and fetus. Thus, it is uniquely situated to evaluate prenatal exposures in the context of the DOHaD hypothesis. Additionally, in 2022, Bhattacharya and colleagues demonstrated the profound health impact of placental genomic regulation in early life traits, that may persist later in life as etiologic antecedents for complex traits, therefore programming development across the life course. In the last years, placental omics, including epigenomic data that are partly regulated by genetics, have been proven useful to study the numerous relationships between prenatal exposure and fetal and early life health outcomes.

2. Epigenetics, the bridge between genetics and environmental factors

DNA methylation (DNAm) is an epigenetic process that regulates gene expression and has been considered a bridge between the environment and the genome, that at the same time, is under the control of both environmental and genetic factors.

It has been observed that DNAm undergoes extensive alterations not only throughout gestation, but also in response to *in utero* conditions. Nowadays, there are many studies proposing that the intrauterine environment alters placental function trough DNAm. It is important to note that, while DNAm is bimodally distributed in most tissues and cell types, with most of the CpG sites either hypomethylated or hypermethylated, placental DNAm follows a trimodal distribution, mostly due to its high content of both partially methylated domains (PMD) and CpG positions with intermediate methylation levels.

However, the particularities of placental DNAm are not only observed in bulk tissue, but also in placental cell types. In 2021, Yuan and colleagues observed that the most distinct DNAm profiles were those of placental trophoblasts (TB), that during gestation differentiate into syncytiotrophoblasts (STB). STBs are the most common cell type in term placentas and form a continuous, multinucleated, and specialized layer of epithelial cells that orchestrates the biomolecular interaction between mother and fetus.

Few studies or none have specifically investigated the role of cell type-specific epigenome in response to pregnancy exposures and/or leading to diverse health outcomes. The important role of placenta and the unique DNAm profile observed at both bulk tissue and cell type levels, lead us to speculate that the genome-environment interaction and its impact on DNAm is highly likely unique in placenta and deserves further investigation.

2.1. Epigenome-wide association studies

The aim of epigenome-wide association studies (EWAS) is to examine the genome-wide epigenetic variation, predominantly DNAm at CpG sites, to detect differentially

methylated positions (DMP) and regions (DMR) that are significantly associated with phenotypes of interest.

2.1.1. Pregnancy And Childhood Epigenetics consortium studies

In the last several years, the Pregnancy And Childhood Epigenetics (PACE) consortium from the National Institute of Health (NIH) has led several EWAS that demonstrate that peripheral and cord blood DNAm is sensitive to environmental factors surrounding gestation. But, regarding placental DNAm, a smaller number of studies have examined the effect of prenatal exposures in this organ, and the ones that did it have demonstrated the tissue-specificity of the DNAm signatures.

Besides, maternal pre-pregnancy body mass index (ppBMI) is of particular interest since higher maternal ppBMI is associated with aberrant fetal growth, macrosomia and increased neonatal morbidity, as well as with pregnancy complications, including preeclampsia, gestational diabetes, gestational hypertension, pre-term delivery and caesarean section. It has been observed that maternal ppBMI is associated with other offspring health outcomes in later life, including increased risk for obesity and poorer cognitive performance in children.

In 2017, the PACE consortium studied the effect of maternal ppBMI cord blood, showing that this exposure is widely associated with the DNAm at this particular tissue. However, the authors observed that many significant epigenetic effects were modest, and they did not detect enrichments for any particular biological pathway, leaving open questions regarding the potential intra-uterine mechanisms that could be affecting the epigenetic profile of the newborn. In this context, while the epigenetic alternations in cord blood and peripheral blood have been thoroughly investigated, the potential impact of maternal ppBMI in placental DNAm remain poorly explored, and the studies performed in this tissue yielded a limited number of significant results, probably because of their limited sample sizes.

2.2. Methylation quantitative trait loci

Methylation quantitative trait loci (mQTL) are single-nucleotide polymorphisms (SNPs) in which genetic variation is associated with the DNAm levels of a CpG

dinucleotide. In general, QTLs have been used to understand the functional effects of genetic variants, many of which are located in non-coding regions, and associated with numerous traits and diseases. In 2021, Oliva *et al*. characterized the genome-wide DNAm profile of diverse, healthy human tissue types, and demonstrated that mQTLs can reveal a substantial number of molecular links to traits otherwise missed by eQTL approaches, pinpointing putative candidate genes.

In this context and as we have already mentioned, DNAm profiles are not only tissue-specific, but also cell type-specific. Therefore, Kim-Hellmuth and colleagues found that cell type-interacting QTLs (iQTL) are strongly enriched for tissue and cellular specificity and provide a finer resolution to tissue specificity than that of bulk QTLs. Thus, iQTLs are highly valuable instruments that allow us to gain mechanistic understanding of complex trait associations.

2.2.1. Placental mQTL studies

In the last few years, there have been several studies that have mapped placental mQTLs such as Do *et al.* 2016, Delahaye *et al.* 2018, Tekola-Ayele *et al.* 2022 and Casazza *et al.* 2024. However, there is neither a publicly accessible placental mQTL database measuring DNAm on more CpG sites other than those on the Illumina 450K array, nor one with a considerable sample size, or one taking into account cell type heterogeneity and its specific epigenetic profiles.

3. Mendelian Randomization and Colocalization analyses

Nowadays, genome-wide association studies (GWAS) have identified thousands of genetic variants associated with human complex traits. However, the genes or functional DNA elements through which these variants exert their effects on the traits remain unknown. Inferring causal relationships between phenotypes is a major challenge and has important implications for understanding the etiology of health and disease. Approaches such as colocalization and mendelian randomization (MR) analyses can be useful to statistically infer the causal effects between the two traits, namely a molecular trait (i.e. DNAm) and a disease.

Colocalization analyses test whether two independent association signals at the same locus are consistent with a shared causal variant. If the answer is positive, we refer to this situation as colocalized traits, and the probability that both traits share a causal mechanism greatly increases. A common example involves an eQTL study and a disease association result, which point to the causal gene and the tissue in which it exerts its effect on the disease.

Besides, MR uses genetic variation to address causal questions about how modifiable exposures influence different outcomes. The principles of MR are based on Mendel's laws of inheritance and instrumental variable estimation methods, which enable the inference of causal effects in the presence of unobserved confounding. Briefly, MR methodology can be interpreted as an analysis to test if the effect of a SNP on an outcome, such as a disease, is mediated by an exposure, such as DNAm.

4. The association between neuropsychiatric traits and the prenatal environment

Neuropsychiatric conditions that usually do not manifest until childhood or young adulthood are increasingly being recognized to have their origins in the fetal period. As such, maternal health and intrauterine environmental exposures can influence fetal brain development, often through converging physiological pathways. Thus, they had been described as significant risk factors for the development of common disorders such as autism spectrum disorder (ASD), attention-deficit and hyperactivity disorder (ADHD), schizophrenia (SCZ) and bipolar disorder (BIP). For instance, other neuropsychiatric-related outputs studied in this context are behavioral outcomes such as internalizing (INT) and externalizing behaviors, which include major depression disorder (MDD), obsessive-compulsive disorder (OCD) and aggressive behavior in children (AGR).

In conclusion, more and more studies are demonstrating the association between pregnancy exposures, such as maternal ppBMI, and the later development of a wide range of traits and disorders in the context of the DOHaD hypothesis. Similarly, epigenetic studies have demonstrated to function as a bridge between genetics and

environmental factors, but most of them have been carried in peripheral and cord blood tissue, leaving aside the potential role of placenta as one of the key players in a successful pregnancy. In the last few years, novel studies have shown that this organ has specific and unique epigenetics marks, different from those identified in cord blood or any other tissue. These patterns might be one of the mechanisms by which prenatal exposures and susceptibility genes exert their effects. In this sense, some neuropsychiatric traits have been suggested to have prenatal origins, but a limited number of articles have focused on the placental epigenome as the potential mechanism by which genetics might be contributing to the development of these disorders.

Aims

The overall aim of this thesis is to elucidate the impact of maternal ppBMI, a relevant fetal environmental factor, and fetal genetics on later life health via placental DNAm.

The operative aims of the present study were:

1. To determine the possible association of maternal ppBMI with epigenome-wide placental DNAm, and to understand the mechanisms by which maternal obesity could be associated with future health outcomes in offspring.
2. To construct and to make publicly available the largest placental mQTL database to date, as a novel tool to understand the effect of placental DNAm in later life health.
    a. To identify and to characterize the placental DNAm sites regulated by genetics, and more specifically by SNPs.
    b. To identify and to characterize the placental DNAm sites regulated by the interaction between SNPs and cell type proportions, gestational age (GA), and fetal sex.
3. To ascertain whether placental DNAm is the mechanism by which genetic risk factors contribute to the development of several neuropsychiatric diseases,

including ADHD, AGR, ASD, BIP, INT, MDD OCD, panic disorder (PD), suicidal attempt (SA), and SCZ.

Methods

1. Placenta biopsies and DNA extraction

The Environment and Childhood study (INMA – *Infancia y Medio Ambiente*) is a population-based mother-child cohort study in Spain that aims to study the role of environmental pollutants in air, water and diet during pregnancy and early childhood in relation to child growth and development. In INMA, 2,506 mother fetus pairs from Asturias, Gipuzkoa, Sabadell and Valencia were followed until birth and a selection of 489 placentas were collected. Biopsies from the inner region of the placenta were obtained, and its DNA extracted and isolated.

2. Genotype data quality control

The genome-wide genotyping of 509,450 genetic variants in 397 INMA samples was performed with the Illumina GSA BeadChip. The quality control of the genotype was performed with PLINK 1.9 software following standard recommendations. The genotype was imputed with the Michigan Imputation Server using the Haplotype Reference Consortium panel, and variants with an imputation quality r^2 below 0.9, a minor allele frequency lower than 5%, a Hardy-Weinberg Equilibrium P value below 0.5 and with more than two alleles were removed. Only those samples with methylation data were considered in this analysis. The final data set consisted of 368 samples and 4,171,035 SNPs.

3. DNAm data quality control

Following the manufacturer's protocol, 190 and 397 placentas were analyzed with the Illumina 450K and EPIC arrays for the maternal ppBMI EWAS and the placental mQTL database, respectively. Briefly, DNAm data was normalized, and low-quality

samples and conflicting probes were discarded. In the case of the maternal ppBMI, the final data set consisted of 168 samples and 405,301 DNAm probes or CpG sites. And, in the case of the placental mQTLs, the final data set consisted of 368 samples and 747,486 DNAm probes or CpG sites.

Additionally, the cellular heterogeneity of the samples was estimated with RefFreeCellMix and Planet R packages for maternal ppBMI and placental mQTLs studies, respectively.

4. EWAS of the maternal ppBMI and placental DNAm

Robust linear regressions were run to test the associations between placental DNAm beta values at each CpG site and maternal ppBMI as a continuous variable. Models were adjusted for maternal age, parity, maternal education and maternal smoking during pregnancy. Two EWAS models were run with and without adjustment for RefFreeCellMix putative cellular components.

The meta-analysis was performed with 2,631 samples from ten PACE cohorts (AQUA, EARLI, EDEN, Gen3G, GENEIDA, HEBC, INMA, ITU, MARBLES, NHBCS and RICHS), whose DNAm quality control protocols were the same as the one described above for the INMA cohort.

Finally, maternal ppBMI-sensitive CpG sites were tested for functional enrichment, as well as for proximity with SNPs associated with birth outcomes and with maternal ppBMI-associated CpG sites reported in a previous PACE cord blood study.

5. Placental *cis*-mQTL analysis

Linear regressions between the genotype and the DNAm rank-based normal transformed-values were performed. The model was adjusted with the sex of the fetus, genetic and non-genetic DNAm principal components (PC), and the cell type proportions. The window for the mQTL mapping was specified as ±0.5 Mb from each tested CpG site position.

Afterwards, we performed several enrichment tests with functional annotations,

such as tissue-specific chromatin and histone marks, as well as gene sets from the Gene Ontology and Disease Ontology databases.

6. Placental *cis*-imQTL analysis

The cell type-interacting mQTLs (imQTL) were computed for STB and TB cell types, and consisted in linear models that included the genotype and one cell type at a time as the interaction term, adjusted by the sex of the fetus and the genetic and non-genetic DNAm PCs. Additionally, we also computed sex-imQTLs and GA-imQTLs correcting by the same covariates plus the cell type estimations. In the case of the sex-imQTLs, the sex of the fetus was excluded as covariate.

7. Multi-omics approaches to unravel the potential placental origin of some neuropsychiatric disorders

The GWAS summary-statistics used in these analyses were the largest, publicly available association studies of ADHD, AGR, ASD, BIP, INT, MDD, OCD, PD, SA and SCZ. These summary-statistics were harmonized to dbSNP build 155 with INMA genotype data as a reference.

MR analyses were carried out considering CpG DNAm as the exposure and the neuropsychiatric disorders as the outcome, thus testing for the causative effect of placental DNAm on these traits. Additionally, a heterogeneity test was applied to discard linkage scenarios where two different SNPs in very strong linkage disequilibrium are affecting the exposure and the outcome separately. The same MR analysis was applied to imQTL databases, but in the case of STB- and TB-imQTLs, only those showing an increase of the main genotype effect with the cell type abundance were considered.

Colocalization analysis was conducted between the significant loci defined in the original GWAS and the placental mQTLs. This analysis was only performed with the three phenotypes studied in more detail (BIP, MDD and SCZ).

Finally, we evaluated the influence of the identified CpG sites on placental gene

expression in an expression quantitative trait methylation (eQTM) database from the North American RICHS cohort. The gene set enrichment analysis of the eQTM-genes associated to SCZ was performed using the Reactome pathway database and website tool.

8. Conditional analysis between placental *cis*-mQTLs and neuropsychiatric disorders

We performed a conditional analysis conditioning on the top mQTL of those probes that passed MR but failed to pass the heterogeneity test, due to the fact that heterogeneity may sometimes be driven by real secondary signals. This conditional analysis was only performed with the three phenotypes studied in more detail (BIP, MDD and SCZ).

Results

1. Maternal ppBMI is associated to 27 placental DNAm sites

After applying the Bonferroni correction for multiple-testing, we obtained 27 and 42 CpG sites at which maternal ppBMI was significantly associated with placental DNAm levels in the models adjusted and unadjusted for cell type components, respectively. High maternal ppBMI was associated with lower placental DNAm in 24/27 differentially methylated CpG sites identified in the cell type adjusted model, while in the unadjusted model, 33/42 hits showed positive associations. However, the effect-size of CpG sites that were differentially methylated in one model were positively correlated to the effect sizes of the same position in the other model. Finally, the heterogeneity of associations across cohorts was lower for the model adjusted for cell type proportions compared to the unadjusted model and thus, we continued with the results from the fully adjusted model for downstream analyses.

Among the 27 DMPs identified in our cell type-adjusted EWAS, a few individual CpG

sites are worthy of mention. For instance, the most significant hit, cg08219219, showed that a 10-unit increase in maternal ppBMI is associated with a 1.1% lower DNAm at this specific CpG site. The largest positive effect size was observed in cg14704941, with 2% higher placental DNAm per 10-unit BMI. In turn, the largest negative effect size was found in cg04724807 with 1.8% lower DNAm per 10-unit BMI.

The 26 genes annotated to the 27 maternal ppBMI-sensitive CpG sites presented an enrichment on small cell lung cancer and oxidative stress-induced signaling pathway gene sets. Additionally, this same list of genes was enriched for genes regulated by the *ZNF217* transcriptional factor.

We also assessed whether the maternal ppBMI-sensitive CpG sites were close to SNPs associated to birth outcomes, and if the DNAm signature of maternal ppBMI in placenta was consistent with associations in cord blood previously reported by the PACE consortium. On the one hand, more than a third of the 27 ppBMI associated CpG sites were within ±0.5 Mb of birthweight SNPs. And, on the other hand, we did not find any overlapping CpG sites associated with maternal ppBMI between the two tissues.

2. More than 200,000 placental CpG sites are regulated by genetic variants in *cis*

We detected 214,830 placental mQTLs based on 214,830 and 150,649 mSites (CpG sites) and mVariants (SNPs), respectively. Most of them presented the mVariants and mSites close to each other, indicating that genetically modulated DNAm is typically close to the implicated regulatory variant.

In general, mSites were depleted from CpG islands and promoters, and enriched within gene bodies and genomic features with intermediate methylation values such as open sea, and CpG island shelf and shore regions. Additionally, there was an enrichment of the human leukocyte antigen region, as well as placenta-specific open chromatin and transcriptionally active regions. The genes close to the mSites were enriched in 219 gene sets, including neuropathy, mood disorder, demyelinating disease, and BIP.

The placental mQTL database can be found in the following address: https://irlab.shinyapps.io/shiny_mqtl_placenta/.

3. Cell type-interacting mQTLs are detected for STB and TB cells

In the case of the imQTLs, we obtained 38 and 1 STB- and TB-imQTLs, respectively, with no significant GA- and sex-imQTLs. The higher amount of STB-imQTLs revealed a higher statistical power for the most abundant cell type. The sharing between STB- and TB-imQTLs was notable, with a modest impact of GA. This suggests that the TB-to-STB transition greatly impacts the placental methylome, and the modest impact of GA is at least partially independent.

4. Multi-omics-approaches unravel the potential placental origin of some neuropsychiatric disorders

With MR, no significant hits were found for many of the traits, such as OCD or SA. In turn, DNAm sites were identified for ADHD (n = 1), ASD (n = 1), BIP (n = 30), MDD (n = 28), and particularly SCZ (n = 214). To replicate the pleiotropic associations identified in the MR analyses, we performed colocalization across 64, 102, and 287 genomic regions defined in the GWAS to be associated with BIP, MDD and SCZ, respectively. Regarding mQTLs, from the total amount of 214,830 mSites with FDR < 0.05, 12,228, 10,343 and 38,412 mSites were considered for BIP, MDD and SCZ, respectively. In BIP, 7,123 region-CpG pairs, involving 6,487 DNAm sites and 64 regions, were supportive of colocalization. In MDD, 3,895 region-CpG pairs, involving 3,850 DNAm sites and 79 regions, were supportive of colocalization. Finally, in SCZ, 21,071 region-CpG pairs, involving 19,661 DNAm sites and 274 regions, were supportive of colocalization. When the overlap with MR hits was assessed, out of the 30 MR hits in BIP, 28 DNAm sites (93%) showed evidence of colocalization with BIP; in MDD, 20 DNAm sites (71%) out of the 28 MR hits showed evidence of colocalization; and from the 214 MR hits in SCZ, 198 DNAm sites (92%) colocalized with SCZ.

When the eQTMs form the RICHS cohort were taken into consideration, out of the 28 and 214 MR-significant DNAm sites in MDD and SCZ, we found that 3 and 43 correlated with the expression levels of 4 and 26 significant eQTM-genes, respectively, supporting the placental function of the CpG sites identified.

Next, we identified the intersect between MR, colocalization and eQTM results for each trait. In the case of MDD, one placental DNAm site colocalized with an MDD-associated genomic locus on chromosome 14, and DNAm levels correlated with the placental expression of *LRFN5*, a gene expressed in TB stem cells that could play an important role in placenta. In SCZ, 40 placental DNAm sites colocalized with 12 SCZ-associated loci on chromosomes 1, 3, 6, 7, 8, 11, 12, 18, 19 and 22, with DNAm levels that correlated with the placental expression of 23 nearby genes, among the total 26 eQTM genes, namely *GLB1L3, H2BC6, H3C4, KCNG2, LETM2, NAGA, PGBD1, PSMG3, RNF39, SDCCAG8, SFMBT1, SLC6A16, TOB2P1, TRIM27, TXNL4A, VARS2, VPS37B, ZKSCAN4, ZNF165, ZNRD1-AS1, ZSCAN12, ZSCAN12P1,* and *ZSCAN23*.

Some of those genes are worthy of mention such as the histone coding genes *H2BC6* and *H3C4*. On the one hand, *H2BC6* is a histone that has been identified in the placenta of mothers with unexplained recurrent spontaneous abortions. On the other, *H3C4* is dysregulated in mice embryos during alcohol exposure in early neurulation due to aberrant changes in DNAm patterns. Finally, *VPS37B* is a crucial gene in viral budding. Altogether, those genes take part in immune homeostasis, in viral infection and seem to be relevant to the development of pregnancy.

Finally, we performed a Reactome gene set analysis of the 26 genes at the intersection between MR and eQTM results in SCZ. We observed enrichment for 101 gene sets, including many epigenetic regulatory pathways, suggesting that SCZ risk alleles could change the epigenetic landscape of placenta through the regulation of, among others, the aforementioned histone-coding genes. Additionally, we observed a remarkable enrichment in immune-related pathways such as human cytomegalovirus infection and overall viral infection, reinforcing a hypothesis that points to maternal immune activation (MIA) as a link between placenta and SCZ risk.

5. A secondary potentially causal signal is detected in SCZ

With the conditional analysis, we discovered a region in SCZ, in which two independently associated SNPs had a pleiotropic effect on both cg27314558 and SCZ. This CpG site is surrounded by multiple transfer RNAs (tRNA), an alanine tRNA being the closest. The impaired placental transfer of tRNAs is known to be associated with fetal growth restriction.

6. Detection of pleiotropic associations between BIP and SCZ, and STB- and TB-interacting mQTLs

The MR test between the cell type-imQTLs and BIP, MDD and SCZ, gave one STB-imQTL pleiotropically associated to BIP, and two TB-imQTLs pleiotropically associated to BIP and MDD (one each). For instance, the CpG involved in the TB-imQTL pleiotropically associated with BIP is located in the gene body of *SMAD3*. The overexpression of this gene in placenta has been shown to activate the ability of TBs to form endothelial-like networks, and to be associated with preeclampsia. Additionally, this gene is a well-known target for lithium treatment in BIP.

Discussion

1. Maternal ppBMI EWAS

Here we present the largest EWAS with placental DNAm performed to date. We identify 27 CpG sites at which we observe placental DNAm variations of 0.5–2.0% by 10-unit maternal ppBMI difference. Additionally, our DNAm findings seem to be placenta-specific, showing minimal overlap with a previous meta-analysis in cord blood DNAm in relation to maternal ppBMI. The differentially methylated CpG sites are mainly located in open sea regions, with a complete depletion from CpG islands, and enriched in cancer- and oxidative stress-related pathways.

These observations, together with the fact that maternal ppBMI is associated with placental DNAm at CpG sites located close to obesity-related genes, leads us to

hypothesize that placental DNAm could be one of the mechanisms by which maternal obesity is associated with aberrant fetal growth and maybe, other metabolic health outcomes in offspring later in life. However, we cannot rule out that the changes observed could be just markers of exposure to high ppBMI, without causal effects on later life health, and therefore, our findings will need to be supplemented by functional studies or causal inference analyses to better understand if they truly have a role in pregnancy complications or long-term metabolic outcomes.

2. Placental mQTLs

In conclusion, we find placental mQTLs to be highly placenta-specific, with a remarkable enrichment in genomic regions active in placenta and neurodevelopment- and mental health-related pathways. We prove that they are useful to map the etiologic window of neuropsychiatric disorders to prenatal stages and conclude that part of the genetic risk of BIP, MDD and in particular SCZ, act through placental DNAm at specific genomic loci. In fact, some of the observed associations might be causal rather than pleiotropic due to the presence of secondary association signals in conditional analyses, involvement of cell type-specific imQTLs and association with the expression levels of relevant genes in placenta. It is of particular interest that SCZ-associated placental DNAm correlates with the expression of immune-related and epigenetic-regulatory genes, providing further support to the hypothesis of MIA being a key factor in the prenatal origins of SCZ. Future work, including functional approaches such as *in vitro* modulation of expression in potential effector cell lines, as well as long term follow up in prospective studies, will be needed in order to better characterize our findings and their implication in disease development.

Final remarks

Early life exposures and risk genetics seem to have an impact on the future health of the offspring by modifying placental DNAm. The placental signatures described here have been demonstrated to be, at least partially, unique for this organ, reinforcing its role in the context of the DOHaD hypothesis. This work also highlights the importance

of conducting previous PACE consortium studies in placenta, given the organ-specific signatures identified. Finally, we have pointed to several potentially causal genes that, according to different pieces of evidence, could underlie the pathogenesis of different neuropsychiatric disorders by exerting their effects in placenta. Among others, the intersection of pleiotropy with placental eQTMs, the regional pleiotropy reported, and even the presence of secondary association signals derived from conditional analyses, strongly suggests that part of the genetic risk to suffer from BIP, MDD and very especially SCZ, acts through placental DNAm at specific genomic loci. Future perspectives need to include studies with larger sample sizes and participants from other populations to increase statistical power and to stop the systematic exclusion of non-European descendants.

Conclusions

The conclusions of the doctoral thesis are:

1. Maternal ppBMI impacts placental DNAm. In this thesis we have identified 27 placenta-specific DNAm sites associated to maternal ppBMI, many of which are located in open sea regions, close to obesity-related genes such as *GPX1* and *LGR4* and altogether, enriched in cancer and stress oxidative pathways. We propose that placental DNAm could be one of the mechanisms by which maternal obesity is associated with metabolic health outcomes in newborns and children.
2. We have made publicly available a catalog of placental mQTLs, including 214,830 unique DNAm sites modelled as a function of the genotype of the fetuses. Additionally, we have found that cell type proportions affect placental DNAm and have identified 38 and 1 placental DNAm sites modelled as a function of the proportions of STB and TB, respectively, in a genotype-dependent manner.
    a. The placental *cis*-mQTLs are depleted on usually hypomethylated regions, and in turn, enriched in regions with intermediate DNAm values, such as CpG islands shelves and shores. Moreover, the *cis*-

mQTLs seem highly placenta-specific and interestingly, are enriched in placenta-specific active chromatin marks, and in neuropathy and mood disorder-related pathways, among others.

b. We observe higher statistical power for the most abundant cell type and therefore, a larger amount of STB-imQTLs compared to TB-imQTLs. However, there is a remarkable sharing and, as expected, the allelic effects in the overlapping imQTLs are negatively correlated between the two cell types. The moderate sharing signal between STB- and GA-imQTLs suggests that the scarce effect of GA on the genetic regulation of placental DNAm is at least partially independent of the TB-to-STB transition.

3. Part of the genetic susceptibility to suffer BIP, MDD and particularly SCZ, acts through placental DNAm. The potential causality of several of the observed associations is reinforced by secondary association signals identified in the conditional analysis, the involvement of cell type-imQTLs, and the correlation of the DNAm sites identified with the expression levels of relevant genes in placenta. Additionally, the presence of immune- and epigenetic-related genes, including *VPS37B, H2BC6* and *H3C4*, as eQTMs of the pleiotropically associated CpG sites, reinforce the prenatal origins of SCZ and the role of the MIA in it. In conclusion, the genetic risk of several neuropsychiatric disorders could operate through placental DNAm and associated gene expression in placenta.

BIBLIOGRAPHY

Bibliography

1. Hales, C. N. et al. Fetal and infant growth and impaired glucose tolerance at age 64. *BMJ (Clinical research ed.)* **303**, 1019–1022 (1991).

2. Barker D. J. The origins of the developmental origins theory. *Journal of internal medicine* **261**, 412–417 (2007).

3. Kunz, L. H., & King, J. C. Impact of maternal nutrition and metabolism on health of the offspring. *Seminars in fetal & neonatal medicine* **12**, 71–77 (2007).

4. Cosin-Tomas, M. et al. Prenatal Maternal Smoke, DNA Methylation, and Multi-omics of Tissues and Child Health. *Current environmental health reports* **9**, 502–512 (2022).

5. Scott, S., & Sher, J. Effect of alcohol during pregnancy: a public health issue. *The Lancet. Public health* **8**, e4–e5 (2023).

6. Popova, S. et al. Maternal alcohol use, adverse neonatal outcomes and pregnancy complications in British Columbia, Canada: a population-based study. *BMC pregnancy and childbirth* **21**, 74 (2021).

7. Mulder, E. J. et al. Prenatal maternal stress: effects on pregnancy and the (unborn) child. *Early human development* **70**, 3–14 (2002).

8. Kwok, J. et al. Maternal infections during pregnancy and child cognitive outcomes. *BMC pregnancy and childbirth* **22**, 848 (2022).

9. Ketema, D. B. et al. Effects of maternal education on birth preparedness and complication readiness among Ethiopian pregnant women: a systematic review and meta-analysis. *BMC pregnancy and childbirth* **20**, 149 (2020).

10. Monk, C., & Fernández, C. R. Neuroscience Advances and the Developmental Origins of Health and Disease Research. *JAMA network open* **5**, e229251 (2022).

11. Walker, C. L., & Ho, S. M. Developmental reprogramming of cancer susceptibility. *Nature Reviews Cancer* **12**, 479–486 (2012).

12. Lu, C. et al. Interaction effect of prenatal and postnatal exposure to ambient air pollution and temperature on childhood asthma. *Environment international* **167**, 107456 (2022).

13. Whitaker R. C. Predicting preschooler obesity at birth: the role of maternal obesity in early pregnancy. *Pediatrics* **114**, e29–e36 (2004).

14. Chen, T. et al. Developmental origins of inflammatory and immune diseases. *Molecular human reproduction* **22**, 858–865 (2016).

15. Alexander, B. T. et al. Fetal programming and cardiovascular pathology. *Comprehensive Physiology* **5**, 997–1025 (2015).

16. O'Donnell, K. et al. Prenatal stress and neurodevelopment of the child: focus on the HPA axis and role of the placenta. *Developmental neuroscience* **31**, 285–292 (2009).

17. Myatt L. Placental adaptive responses and fetal programming. *The Journal of physiology* **572**, 25–30 (2006).

18. Saenen, N. D. et al. Air pollution-induced placental alterations: an interplay of oxidative stress, epigenetics, and the aging phenotype? *Clinical epigenetics*, **11**, 124 (2019).

19. Lapehn, S. & Paquette, A.G. The Placental Epigenome as a Molecular Link Between Prenatal Exposures and Fetal Health Outcomes Through the DOHaD Hypothesis. *Current environmental health reports* **9**, 490-501 (2022).

20. Mortillo, M., & Marsit, C. J. Select Early-Life Environmental Exposures and DNA Methylation in the Placenta. *Current environmental health reports* **10**, 22–34 (2023).

21. Wang, Y., & Zhao, S. (2010). *Vascular Biology of the Placenta*. *Morgan & Claypool Life Sciences*. https://doi.org/10.4199/C00016ED1V01Y201008ISP009

22. Bhattacharya, A. et al. Placental genomics mediates genetic associations with complex health traits and disease. *Nature Communications* **13**, 706 (2022).

23. Peng, S. et al. Genetic regulation of the placental transcriptome underlies birth weight and risk of childhood obesity. *PLOS Genetics* **14**, e1007799 (2018).

24. Shallie, P. D., & Naicker, T. The placenta as a window to the brain: A review on the role of placental markers in prenatal programming of neurodevelopment. *International journal of developmental neuroscience: the official journal of the International Society for Developmental Neuroscience* **73**, 41–49 (2019).

25. Behura, S. K. et al. The brain-placental axis: Therapeutic and pharmacological relevancy to pregnancy. *Pharmacological research* **149**, 10446 (2019).

26. Tedner, S. G. et al. Fetal growth and risk of childhood asthma and allergic disease. *Clinical and experimental allergy: journal of the British Society for Allergy and Clinical Immunology* **42**, 1430–1447 (2012).

27. Everson, T. M., & Marsit, C. J. Integrating -Omics Approaches into Human Population-Based Studies of Prenatal and Early-Life Exposures. *Current environmental health reports* **5**, 328–337 (2018).

28. Weinhold B. Epigenetics: the science of change. *Environmental health perspectives* **114**, A160–A167 (2006).

29. Handy, D. E. et al. Epigenetic modifications: basic mechanisms and role in cardiovascular disease. *Circulation* **123**, 2145–2156 (2011).

30. Moore, L. D., et a. DNA methylation and its basic function. *Neuropsychopharmacology: official publication of the American College of Neuropsychopharmacology* **38**, 23–38 (2013).

31. Bird, A. et al. A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell* **40**, 91–99 (1985).

32. Larsen, F. et al. CpG islands as gene markers in the human genome. *Genomics* **13**, 1095–1107 (1992).

33. Frigola, J. et al. Differential DNA hypermethylation and hypomethylation signatures in colorectal cancer. *Human molecular genetics* **14**, 319–326 (2005).

34. Cross, S. H., & Bird, A. P. CpG islands and genes. *Current opinion in genetics & development* **5**, 309–314 (1995).

35. Holliday, R., & Pugh, J. E. DNA modification mechanisms and gene activity during development. *Science* **187**, 226–232 (1975).

36. Meissner, A. et al. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454**, 766–770 (2008).

37. Hogg, K. et al. Prenatal and perinatal environmental influences on the human fetal and placental epigenome. *Clinical pharmacology and therapeutics* **92**, 716–726 (2012).

38. Novakovic, B. & Saffery, R. The ever growing complexity of placental epigenetics - role in adverse pregnancy outcomes and fetal programming. *Placenta* **33**, 959-970 (2012).

39. Konwar, C. et al. Considerations when processing and interpreting genomics data of the placenta. *Placenta* **84**, 57-62 (2019).

40. Schroeder, D. I. et al. The human placenta methylome. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 6037–6042 (2013).

41. Schmidt, R. J. et al. Self-reported pregnancy exposures and placental DNA methylation in the MARBLES prospective autism sibling study. *Environmental epigenetics* **2**, dvw024 (2016).

42. Yuan, V. et al. Cell-specific characterization of the placental methylome. *BMC genomics*, **22**, 6 (2021).

43. Mayhew, T. M. & Barker, B.L. Villous trophoblast: morphometric perspectives on growth, differentiation, turnover and deposition of fibrin-type fibrinoid during gestation. *Placenta* **22**, 628-638 (2001).

44. Ellery, P. M. et al. Evidence for transcriptional activity in the syncytiotrophoblast of the human placenta. *Placenta* **30**, 329-334 (2009).

45. White, R. et al. Environmentally persistent alkylphenolic compounds are estrogenic. *Endocrinology* **135**, 175–182 (1994).

46. McLaren, A. Early embryo-endometrial relationships. Ovo-Implantation. Human Gonadotropins and Prolactin. *Karger* 18–73 (1970).

47. Finn, C. A., & Bredl, J. C. Studies on the development of the implantation reaction in the mouse uterus: influence of actinomycin D. *Journal of reproduction and fertility* **34**, 247–253 (1973).

48. Braga, V. M., & Gendler, S. J. Modulation of Muc-1 mucin expression in the mouse uterus during the estrus cycle, early pregnancy and placentation. *Journal of cell science* **105**, 397–405 (1993).

49. Wang, M. et al. Single-nucleus multi-omic profiling of human placental syncytiotrophoblasts identifies cellular trajectories during pregnancy. *Nature Genetics* **56**, 294–305 (2024).

50. Campagna, M. P. et al. Epigenome-wide association studies: current knowledge, strategies and recommendations. *Clinical epigenetics* **13**, 214 (2021).

51. Bibikova, M. et al. High density DNA methylation array with single CpG site resolution. *Genomics* **98**, 288–95 (2011).

52. Moran, S. et al. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. *Epigenomics* **8**, 389–99 (2016).

53. Felix, J. F. et al. Cohort Profile: Pregnancy And Childhood Epigenetics (PACE) Consortium. *International journal of epidemiology* **47**, 22–23 (2018).

54. Joubert, B. R. et al. DNA Methylation in Newborns and Maternal Smoking in Pregnancy: Genome-wide Consortium Meta-analysis. *The American journal of human genetics* **98**, 680–696 (2016).

55. National Institute of Environmental Health Sciences. (2021, September 22). *PACE Projects*. National Institute of Health. https://www.niehs.nih.gov/research/atniehs/labs/epi/pi/genetics/pace/projects

56. Gruzieva, O. et al. Epigenome-Wide Meta-Analysis of Methylation in Children Related to Prenatal NO2 Air Pollution Exposure. *Environmental health perspectives* **125**, 104–110 (2017).

57. Sharp, G. C. et al. Maternal BMI at the start of pregnancy and offspring epigenome-wide DNA methylation: findings from the pregnancy and childhood epigenetics (PACE) consortium. *Human molecular genetics* **26**, 4067–4085 (2017).

58. Sharp, G. C. et al. Maternal alcohol consumption and offspring DNA methylation: findings from six general population-based birth cohorts. *Epigenomics* **10**, 27–42 (2018).

59. Yeung, E. H. et al. Cord blood DNA methylation reflects cord blood C-reactive protein levels but not maternal levels: a longitudinal study and meta-analysis. *Clinical epigenetics* **12**, 60 (2020).

60. Ronkainen, J. et al. Maternal haemoglobin levels in pregnancy and child DNA methylation: a study in the pregnancy and childhood epigenetics consortium. *Epigenetics* **17**, 19–31 (2022).

61. Reese, S. E. et al. Epigenome-wide meta-analysis of DNA methylation and childhood asthma. *The Journal of allergy and clinical immunology* **143**, 2062–2074 (2019).

62. Küpers, L. K. et al. Meta-analysis of epigenome-wide association studies in neonates reveals widespread differential DNA methylation associated with birthweight. *Nature Communications* **10**, 1893 (2019).

63. Vehmeijer, F. O. L. et al. DNA methylation and body mass index from birth to adolescence: meta-analyses of epigenome-wide association studies. *Genome medicine* **12**, 105 (2020).

64. Neumann, A. et al. Association between DNA methylation and ADHD symptoms from birth to school age: a prospective meta-analysis. *Translational psychiatry* **10**, 398 (2020).

65. Everson, T. M. et al. Placental DNA methylation signatures of maternal smoking during pregnancy and potential impacts on fetal growth. *Nature Communications* **12**, 5095 (2021).

66. Maffeis, C. & Morandi, A. Effect of Maternal Obesity on Foetal Growth and Metabolic Health of the Offspring. *Obesity Facts* **10**, 112–117 (2017).

67. Marchi, J. et al. Risks associated with obesity in pregnancy, for the mother and baby: a systematic review of reviews. *Obesity Reviews* **16**, 621–638 (2015).

68. Howell, K. R. & Powell, T. L. Effects of maternal obesity on placental function and fetal development. *Reproduction* **153**, R97–R108 (2017).

69. Mamun, A. A. et al. Gestational weight gain in relation to offspring obesity over the life course: a systematic review and bias-adjusted meta-analysis. *Obesity Reviews* **15**, 338–347 (2014).

70. Godfrey, K. M. et al. Influence of maternal obesity on the long-term health of offspring. *The Lancet Diabetes & Endocrinology* **5**, 53–64 (2017).

71. Rivera, H. M. et al. The role of maternal obesity in the risk of neuropsychiatric disorders. *Frontiers in Neuroscience* **9**, 194 (2015).

72. Håberg S.E. et al. Maternal obesity in pregnancy and respiratory health in early childhood. *Paediatric and Perinatal Epidemiology* **23**, 352–362 (2009).

73. Harskamp-van Ginkel M.W. et al. A study on mediation by offspring BMI in the association between maternal obesity and child respiratory outcomes in the amsterdam born and their development study cohort. *PLOS ONE* **10**, e0140641(2015).

74. Van Lieshout R.J. et al. Pre-pregnancy and pregnancy obesity and neurodevelopmental outcomes in offspring: a systematic review. *Obesity Reviews* **12**, e548–e559 (2011).

75. Sharp, G. C. et al. Maternal pre-pregnancy BMI and gestational weight gain, offspring DNA methylation and later offspring adiposity: findings from the Avon Longitudinal Study of Parents and Children. *International Journal of Epidemiology* **44**, 1288–1304 (2015).

76. Shrestha, D. et al. Placental DNA methylation changes associated with maternal prepregnancy BMI and gestational weight gain. *International Journal of Obesity* **44**, 1406–1416 (2020).

77. Thakali, K. M. et al. Associations between maternal body mass index and diet composition with placental DNA methylation at term. *Placenta* **93**, 74–82 (2020).

78. Zheng, Z. et al. QTLbase: an integrative resource for quantitative trait loci across multiple human molecular phenotypes. *Nucleic acids research*, **48**, D983–D991 (2020).

79. Nicolae, D. L. et al. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLOS Genetics* **6**, e1000888 (2010).

80. Cano-Gamez, E. & Trynka, G. From GWAS to function: Using functional genomics to identify the mechanisms underlying complex diseases. *Frontiers in Genetics* **11**, 424 (2020).

81. GTEx Consortium et al. Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).

82. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).

83. Chen, L. et al. Genetic drivers of epigenetic and transcriptional variation in human immune cells. *Cell* **167**, 1398–1414.e24 (2016).

84. Hannon, E. et al. Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci. *Nature Neuroscience* **19**, 48–54 (2016).

85. Morrow, J. D. et al. Human lung DNA methylation quantitative trait loci colocalize with chronic obstructive pulmonary disease genome-wide association loci. *American Journal of Respiratory and Critical Care Medicine* **197**, 1275–1284 (2018).

86. Taylor, D. L. et al. Integrative analysis of gene expression, DNA methylation, physiological traits, and genetic variation in human skeletal muscle. *PNAS* **116**, 10883–10888 (2019).

87. Andrews, S. V. et al. Cross-tissue integration of genetic and epigenetic data offers insight into autism spectrum disorder. *Nature Communications* **8**, 1–10 (2017).

88. Ng, B. et al. An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. *Nature Neuroscience* **20**, 1418–1426 (2017).

89. Huan, T. et al. Genome-wide identification of DNA methylation QTLs in whole blood highlights pathways for cardiovascular disease. *Nature Communications*. **10**, 4267 (2019).

90. Oliva, M. et al. DNA methylation QTL mapping across diverse human tissues provides molecular links between genetic variation and complex traits. *Nature Genetics* **55**, 112-122 (2023).

91. Kim-Hellmuth, S. et al. Cell type-specific genetic regulation of gene expression across human tissues. *Science* **369**, eaaz8528 (2020).

92. Fairfax, B. P. et al. Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nature Genetics* **44**, 502–510 (2012).

93. Raj, T. et al. Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. *Science* **344**, 519–523 (2014).

94. Naranbhai, V. et al. Genomic modulators of gene expression in human neutrophils. *Nature Communications* **6**, 7545 (2015).

95. Kim-Hellmuth, S. et al. Genetic regulatory effects modified by immune activation contribute to autoimmune disease associations. *Nature Communications* **8**, 266 (2017).

96. Kasela, S. et al. Pathogenic implications for autoimmune mechanisms derived by comparative eQTL analysis of $CD4^+$ versus $CD8^+$ T cells. *PLOS Genetics* **13**, e1006643–e21 (2017).

97. Van der Wijst, M. G. P. et al. Single-cell RNA sequencing identifies cell type-specific cis-eQTLs and co-expression QTLs. *Nature Genetics* **50**, 493–497 (2018).

98. Sarkar, A. K. et al. Discovery and characterization of variance QTLs in human induced pluripotent stem cells. *PLOS Genetics* **15**, e1008045–e16 (2019).

99. Mandric, I. et al. (2020). Optimized design of single-cell RNA sequencing experiments for cell-type-specific eQTL analysis. *Nature Communications* **11**, 5504 (2020).

100. Van der Wijst, M. G. P. et al., Science Forum: The single-cell eQTLGen consortium. *eLife* **9**, e52155 (2020).

101. Westra, H. J. et al. Cell Specific eQTL Analysis without Sorting Cells. *PLOS Genetics* **11**, e1005223–e17 (2015).

102. Zhernakova, D. V. et al. Identification of context-dependent expression quantitative trait loci in whole blood. *Nature Genetics* **49**, 139–145 (2017).

103. Glastonbury, C. A. et al. Cell-Type Heterogeneity in Adipose Tissue Is Associated with Complex Traits and Reveals Disease-Relevant Cell-Specific eQTLs. *The American Journal of Human Genetics* **104**, 1013–1024 (2019).

104. Do, C. et al. Mechanisms and Disease Associations of Haplotype-Dependent Allele-Specific DNA Methylation. *The American Journal of Human Genetics* **98**, 934-955 (2016).

105. Delahaye, F. et al. Genetic variants influence on the placenta regulatory landscape. *PLOS Genetics* **15**, e1008118 (2018).

106. Wang, Y. et al. ZNF217: An Oncogenic Transcription Factor and Potential Therapeutic Target for Multiple Human Cancers. *Cancer management and research*, **16**, 49–62 (2024).

107. Tekola-Ayele, F. et al. Placental multi-omics integration identifies candidate functional genes for birthweight. *Nature Communications* **13**, 2384 (2022).

108. Warrington, N. M. et al. Maternal and fetal genetic effects on birth weight and their relevance to cardio-metabolic risk factors. *Nature Genetics* **51**, 804–814 (2019).

109. Casazza, W. et al. Sex-dependent placental methylation quantitative trait loci provide insight into the prenatal origins of childhood onset traits and conditions. iScience, **27**, 109047 (2024).

110. Nica, A. C. et al. Candidate causal regulatory effects by integration of expression qtls with complex trait genetic associations. *PLOS Genetics* **6**, e1000895 (2010).

111. Hunt, K. A. et al. Newly identified genetic risk variants for celiac disease related to the immune response. *Nature Genetics* **40**, 395–402 (2008).

112. He, X. et al. Sherlock: Detecting gene-disease associations by matching patterns of expression qtl and gwas. *The American Journal of Human Genetics* **92**, 667–680 (2013).

113. Giambartolomei, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLOS Genetics* **10**, e1004383 (2014).

114. Deng, Z. et al. (2024). Identification of TACSTD2 as novel therapeutic targets for cisplatin-induced acute kidney injury by multi-omics data integration. *Human Genetics* (2014).

115. Zhang, H. et al. Altered DNA methylation within DNMT3A, AHRR, LTA/TNF

loci mediates the effect of smoking on inflammatory bowel disease. *Nature Communications* **15**, 595 (2024).

116. Sanderson, E. et al. Mendelian randomization. *Nature Reviews. Methods primers*, **2**, 6 (2022).

117. Hemani, G. et al. The MR-Base platform supports systematic causal inference across the human phenome. *eLife*, **7**, e34408 (2018).

118. Pierce, B.L. & Burgess, S. Efficient design for mendelian randomization studies: Subsample and 2-sample instrumental variable estimators. *American Journal of Epidemiology* **178**, 1177–1184 (2013).

119. Smith, D. & Hemani, G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Human Molecular Genetics* **23**, R89–R98 (2014).

120. Haycock, P. C. et al. Best (but oft-forgotten) practices: the design, analysis, and interpretation of Mendelian randomization studies. *The American Journal of Clinical Nutrition* **103**, 965–978 (2016).

121. Zhu, Z., et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nature Genetics* **48**, 481-487 (2016).

122. Rahmioglu, N. et al. The genetic basis of endometriosis and comorbidity with other pain and inflammatory conditions. *Nature Genetics* **55**, 423–436 (2023).

123. Liu, H. et al. Epigenomic and transcriptomic analyses define core cell types, genes and targetable mechanisms for kidney disease. *Nature Genetics* **54**, 950–962 (2022).

124. Xu, S. et al. Oxidative stress gene expression, DNA methylation, and gut microbiota interaction trigger Crohn's disease: a multi-omics Mendelian randomization study. *BMC medicine* **21**, 179 (2023).

125. Gata-Garcia, A. & Diamond, B. Maternal antibody and ASD: clinical data and animal models. *Frontiers in Immunology* **10**, 1129 (2019).

126. Nielsen, P. R. et al. Association between parental hospital-treated infection and the risk of schizophrenia in adolescence and early adulthood. *Schizophrenia Bulletin* **39**, 230–237 (2013).

127. Jiang, H. Y. et al. Maternal infection during pregnancy and risk of autism spectrum disorders: a systematic review and meta-analysis. *Brain, Behavior, and Immunity* **58**, 165–172 (2016).

128. Zhou, Y. Y. et al. Maternal infection exposure and the risk of psychosis in the offspring: a systematic review and meta-analysis. *Journal of Psychiatric Research* **135**, 28–36 (2021).

129. Meyer, U. et al. In-vivo rodent models for the experimental investigation of prenatal immune activation effects in neurodevelopmental brain disorders. *Neuroscience & Biobehavioral Reviews* **33**, 1061- 1079 (2009).

130. Estes, M. L. & McAllister, A. K. Maternal immune activation: Implications for neuropsychiatric disorders. *Science* **353**, 772-777 (2016).

131. Boksa, P. & El-Khodor, B. F. Birth insult interacts with stress at adulthood to alter dopaminergic function in animal models: possible im-plications for schizophrenia and other disorders. *Neuroscience & Biobehavioral Reviews* **27**, 91-101 (2003).

132. Reinebrant, H. E. et al. Neonatal hypoxia- ischaemia disrupts descending neural inputs to dorsal raphe nuclei. *Neuroscience* **248**, 427-435 (2013).

133. Buller, K. M. et al. Selective losses of brainstem catecholamine neurons after hypoxia- ischemia in the immature rat pup. *Pediatrics Research* **63**, 364-369 (2008).

134. Muller, C. L. et al. The serotonin system in autism spectrum disorder: From biomarker to animal models. *Neuroscience* 321, 24-41(2016).

135. Blows, W. T. Neurotransmitters of the brain: serotonin, nor-adrenaline (norepinephrine), and dopamine. *Journal of Neuroscience Nursing* **32**, 234-238 (2000).

136. Tripp, G. & Wickens, J. R. Neurobiology of ADHD. *Neuropharmacology* **57**, 579-589 (2009).

137. Dufour-Rainfray, D. et al. Fetal exposure to teratogens: evidence of genes in-volved in autism. *Neuroscience Biobehavioral Reviews* **35**, 1254-1265 (2011).

138. Zhang, W. et al. Prenatal depression and infant temperament: the moderating role of placental gene expression. *Infancy* **23**, 211-231 (2018).

139. Lin, Q. et al. Prenatal exposure to environmental tobacco smoke and hyperactivity behavior in Chinese young children. *International Journal of Environmental Research and Public Health* **14**, 10 (2017).

140. Suter, M. A. et al. In utero nicotine expo-sure epigenetically alters fetal chromatin structure and differentially regulates transcription of the glucocorticoid receptor in a rat model. *Birth defects research. Part A, Clinical and molecular teratology* **103**, 583-588 (2015).

141. Warton, F. L. et al. Prenatal methamphetamine exposure is associated with reduced subcortical volumes in neonates. *Neurotoxicology and Teratology* **65**, 51-59 (2018).

142. Warton, F. L. et al. Prenatal methamphetamine exposure is associated with corticostriatal white matter changes in neonates. *Metabolic Brain Disease* **33**, 507-522 (2018).

143. Hepper, P. G. et al. Fetal brain function in response to maternal alcohol consumption: early evidence of damage. *Alcohol Clinical and Experimental Research* **36**, 2168-2175 (2012).

144. Taormina, M. K. MSMA Should Embrace Scientific Cannabis Education. *Missouri Medicine* **117**, 529–530 (2020).

145. Kiblawi, Z. N. et al. The effect of prenatal methamphetamine exposure on attention as assessed by continuous performance tests: results from the Infant Development, Environment, and Lifestyle study. *Journal of Developmental and Behavioral Pediatrics* **34**, 31–37 (2013).

146. Pearson, R. M. et al. Maternal depression during pregnancy and the postnatal period: risks and possible mechanisms for offspring depression at age 18 years. *JAMA Psychiatry* **70**, 1312-1319 (2013).

147. Schulz, K. M. et al. The effects of pre-natal stress on alpha4 beta2 and alpha7 hippocampal nicotinic acetylcholine receptor levels in adult offspring. *Developmental Neurobiology* **73**, 806-814 (2013).

148. Oaks, B. M., et al. Increased risk of preterm delivery with high cortisol during pregnancy is modified by fetal sex: a cohort study. *BMC pregnancy and childbirth* **22**, 727 (2022).

149. Eberle, C. et al. Impact of maternal prenatal stress by glucocorticoids on metabolic and cardiovascular outcomes in their offspring: A systematic scoping review. *PLOS ONE* **16**, e0245386 (2021).

150. Fleck, L. et al. Prenatal stress and child externalizing behavior: effects of maternal perceived stress and cortisol are moderated by child sex. *Child and adolescent psychiatry and mental health*, **17**, 94 (2023).

151. Keller, J. et al. HPA axis in major depression: cortisol, clinical symptomatology and genetic variation predict cognition. *Mol Psychiatry.* **22**, 527-536 (2017).

152. Schlotz, W. & Phillips, D. I. Fetal origins of mental health: evidence and mechanisms. *Brain Behav Immun.* **23**, 905-916 (2009).

153. American Psychiatric Association, DSM-5 Task Force. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5™*. American Psychiatric Publishing, Inc.

154. Walker, C. K. et al. Preeclampsia, placental insufficiency, and autism spectrum disorder or developmental delay. *JAMA Pediatr.* **169**, 154-162 (2015).

155. Bahado-Singh, R. O. et al. Artificial intelligence and placental DNA methylation: newborn prediction and molecular mechanisms of autism in preterm children. *J Matern Fetal Neonatal Med*. **17**, 1–10 (2021a).

156. Bahado-Singh, R. O. et al. Placental DNA methylation changes and the early prediction of autism in full-term newborns. *PLOS ONE* **16**, e0253340 (2021b).

157. Bakulski, K. M. et al. Autism-associated DNA methylation at birth from multiple tissues is enriched for autism genes in the early autism risk longitudinal investigation. *Frontiers in Molecular Neuroscience* **14**, 775390 (2021).

158. Santos, H. P. Jr. et al. Evidence for the placenta-brain axis: multi-omic kernel aggregation predicts intellectual and social impairment in children born extremely preterm. *Molecular Autism* **11**, 97 (2020).

159. Zhu, Y. et al. Placental DNA methylation levels at CYP2E1 and IRS2 are associated with child outcome in a prospective autism study. *Human Molecular Genetics* **28**, 2659–74 (2019).

160. Schroeder, D. I. et al. Placental methylome analysis from a prospective autism study. *Molecular Autism* **7**, 51 (2016).

161. Barton, A. & Fendrik, A. J. Sustained vs. oscillating expressions of Ngn2, Dll1 and Hes1: a model of neural differentiation of embryonic telencephalon. *Journal of Theoretical Biology* **328**, 1–8 (2013).

162. Ramos, C. et al. Two Notch ligands, Dll1 and Jag1, are differently restricted in their range of action to control neurogenesis in the mammalian spinal cord. *PLOS ONE*. **5**, e15515 (2010).

163. Deneault, E. et al. Complete disruption of autism-susceptibility genes by gene editing predominantly reduces functional connectivity of isogenic human neurons. *Stem Cell Reports* **11**, 1211–25 (2018).

164. Nehme, R. et al. Combining NGN2 programming with developmental patterning generates human excitatory neurons with NMDAR-mediated synaptic transmission. *Cell Reports* **23**, 2509–23 (2018).

165. Lin, Y. J. et al. NTF3 is a novel target gene of the transcription factor POU3F2 and is required for neuronal differentiation. *Molecular Neurobiology* **55**, 8403–13 (2018).

166. Van der Ende, J. & Tiemeier, H. The bidirectional pathways between internalizing and externalizing problems and academic performance from 6 to 18 years. *Development and Psychopathology* **28**, 855–67 (2016).

167. Farrington, D. P. Early predictors of adolescent aggression and adult violence. *Violence and Victimcs* **4**, 79–100 (1989).

168. Loth, A. K. et al. Do childhood externalizing disorders predict adult depression? A metaanalysis. *Journal of Abnormal Child Psychology* **42**, 1103–13 (2014).

169. Mathyssek, C. M. et al. Childhood internalizing and externalizing problems predict the onset of clinical panic attacks over adolescence: the TRAILS study. *PLOS ONE* **7**, e51564 (2012).

170. Miettunen, J. et al. Longitudinal associations between childhood and adulthood externalizing and internalizing psychopathology and adolescent substance use. *Psychological Medicine* **44**, 1727–38 (2014).

171. Liu, J. Early health risk factors for violence: conceptualization, evidence, and implications. *Aggression and Violent Behavior* **16**, 63–73 (2011).

172. Murray, C. J. et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease study 2010. *The Lancet* **380**, 2197–223 (2012).

173. Achenbach, T. M. The child behavior profile: i. Boys aged 6–11. *Journal of Consultant and Clinical Psychology* **46**, 478–88 (1978).

174. Liu, J. et al. Childhood internalizing behaviour: analysis and implications. *Journal of Psychiatric and Mental Health Nursing* **18**, 884–94 (2011).

175. Achenbach, T. M. et al. Internalizing/externalizing problems: review and recommendations for clinical and research applications. *Journal of American Academy of Child and Adolescence Psychiatry* **55**, 647–56 (2016).

176. Liu, J. Childhood externalizing behavior: theory and implications. *Journal of Child and Adolescence Psychiatric Nursing* **17**, 93–103 (2014).

177. Van Lieshout, R. J. et al. Maternal pre-pregnancy body mass index and internalizing and externalizing problems in offspring. *The Canadian Journal of Psychiatry* **58**, 151–9 (2013).

178. Sayal, K. et al. Prenatal exposure to binge pattern of alcohol consumption: mental health and learning outcomes at age 11. *European Child & Adolescence Psychiatry* **23**, 891–9 (2014).

179. Murphy, S. K. et al. Maternal infection and stress during pregnancy and depressive symptoms in adolescent offspring. *Psychiatry Research* **257**, 102–10 (2017).

180. Dolan, C. V. et al. Testing causal effects of maternal smoking during pregnancy on offspring's externalizing and internalizing behavior. *Behavior Genetics* **46**, 378–88 (2016).

181. Moylan, S. et al. The impact of maternal smoking during pregnancy on depressive and anxiety behaviors in children: the Norwegian Mother and Child Cohort Study. *BMC Medicine* **13**, 24 (2015).

182. Mac Giollabhui, N. et al. Maternal inflammation during pregnancy and offspring psychiatric symptoms in childhood: timing and sex matter. *Journal of Psychiatric Research* **111**, 96–103 (2019).

183. Sexual and Reproductive Health and Research (SRH). *Global and regional estimates of violence against women.* World Health Organization. https://www.who.int/reproductivehealth/publications/violence/9789241564625/en/.

184. Martinez-Torteya, C. et al. Exposure to intimate partner violence in utero and infant internalizing behaviors: moderation by salivary cortisol-alpha amylase asymmetry. *Early Human Development* **113**, 40–8 (2017).

185. McFarlane, J. et al. Effect of abuse during pregnancy on maternal and child safety and functioning for 24 months after delivery. *Obstetrics & Gynecology* **123**, 839–47 (2014).

186. Nakamura, A. et al. Epigenome-Wide Associations of Placental DNA Methylation and Behavioral and Emotional Difficulties in Children at 3 Years of Age. *International journal of molecular sciences*, **24**, 11772 (2023).

187. Magnus, W. et al. (2023). *Attention Deficit Hyperactivity Disorder.* In StatPearls. StatPearls Publishing.

188. Gustavson, K. et al. Maternal fever during preg-nancy and offspring attention deficit hyperactivity disorder. *Scientific Reports* **9**, 9519 (2019).

189. Ji, Y. et al. Association of cord plasma biomarkers of in utero acetaminophen

exposure with risk of attention-deficit/hyperactivity disorder and autism spectrum disorder in childhood. *JAMA Psychiatry* **77**, 180 (2020).

190. Ystrom, E. et al. Prenatal exposure to ac-etaminophen and risk of ADHD. *Pediatrics* **140**, e20163840 (2017).

191. Polanska, K. et al. Exposure to environmental and lifestyle factors and attention-deficit / hyperactivity disorder in children - a review of epidemiological studies. *International Journal of Occupational Medicine and Environmental Health* **25**, 330-355 (2012).

192. Franz, A. P. et al. Attention-deficit/hyperactivity disorder and very preterm/very low birth weight: A meta-analysis. *Pediatrics* **141**, e20171645 (2018).

193. Miyake, K. et al. DNA methylation of GFI1 as a mediator of the association between prenatal smoking exposure and ADHD symptoms at 6 years: the Hokkaido Study on Environment and Children's Health. *Clinical epigenetics* **13**, 74 (2021).

194. Olstad, E. W. et al. No impact of prenatal paracetamol and folic acid exposure on cord blood DNA methylation in children with attention-deficit/hyperactivity disorder. *Frontiers in genetics* **14**, 1204879 (2023).

195. Patel, K. R. et al. Schizophrenia: overview and treatment options. *P & T: a peer-reviewed journal for formulary management* **39**, 638–645 (2014).

196. Mednick, S. A. et al. Adult schizophrenia following prenatal exposure to an influenza epidemic. *Archives of General Psychiatry* **45**, 189-192 (1988).

197. Rapoport, J.L. et al. The neurodevelopmental model of schizophrenia: update 2005. *Molecular Psychiatry* **10**, 434-449 (2005).

198. Murray, R.M. & Lewis, S.W. Is schizophrenia a neurodevelopmental disorder? *BMJ* **295**, 681-682 (1987).

199. Owen, M.J. & O'Donovan, M.C. Schizophrenia and the neurodevelopmental continuum: evidence from genomics. *World Psychiatry* **16**, 227-235 (2017).

200. Canetta, S.E. & Brown, A.S. PRENATAL INFECTION, MATERNAL IMMUNE ACTIVATION, AND RISK FOR SCHIZOPHRENIA. *Translational Neurosciences* **3**, 320-327 (2012).

201. Cannon, M. et al. Obstetric complications and schizophrenia: historical and meta-analytic review. *American Journal of Psychiatry* **159**, 1080–1092 (2002).

202. Carnac, T. Schizophrenia Hypothesis: Autonomic Nervous System Dysregulation of Fetal and Adult Immune Tolerance. *Frontiers in System Neuroscience* **16**, 844383 (2022).

203. Brown, A. S. et al. Elevated maternal interleukin-8 levels and risk of schizophrenia in adult offspring. *American Journal of Psychiatry* **161**, 889-895 (2004).

204. Shi, J. et al. Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature* **460**, 753-757 (2009).

205. Janicijevic, S.M. et al. Interplay of brain-derived neurotrophic factor and cytokines in schizophrenia. *Serbian Journal of Experimental and Clinical Research* **108**, 110–117 (2018).

206. Sozuguzel, M.D. et al. Female gender specific association of the Reelin (RELN) gene rs7341475 variant with schizophrenia. *Molecular Biology Reports* **46**, 3411–3416 (2019).

207. Nishioka, M. et al. DNA methylation in schizophrenia: Progress and challenges of epigenetic studies. *Genome Medicine* **4**, 96 (2012).

208. Martinez-Jauand, M. et al. Pain sensitivity in fibromyalgia is associated with catechol-O-methyltransferase (COMT) gene. *European Journal of Pain* **17**,16–27 (2013).

209. Iwamoto, K. et al. DNA methylation status of SOX10 correlates with its downregulation and oligodendrocyte dysfunction in schizophrenia. *Journal of Neuroscience* **25**, 5376–5381 (2005).

210. Ursini, G. et al. Convergence of placenta biology and genetic risk for schizophrenia. *Nature Medicine* **24**, 792-801 (2018).

211. Ursini, G. et al. Placental genomic risk scores and early neurodevelopmental outcomes. *PNAS* **118**, e2019789118 (2021).

212. Jain, A. & Mitra, P. (2023). *Bipolar Disorder.* In StatPearls. StatPearls Publishing.

213. Vieta, E. et al. Bipolar disorders. *Nature reviews. Disease primers* **4**, 18008 (2018).

214. Robinson, N. & Bergen, S. E. Environmental Risk Factors for Schizophrenia and Bipolar Disorder and Their Relationship to Genetic Risk: Current Knowledge and Future Directions. *Frontiers in Genetics* **12**, 686666 (2021).

215. Forstner, A. J. et al. Identification of shared risk loci and pathways for bipolar disorder and schizophrenia. *PLOS ONE* **12**, e0171595 (2017).

216. Haukvik, U. K. et al. Pre- and perinatal hypoxia associated with hippocampus/amygdala volume in bipolar disorder. *Psychological medicine* **44**, 975–985 (2014).

217. Guxens, M. et al. Cohort Profile: The INMA—INfancia y Medio Ambiente—(Environment and Childhood) Project. *International Journal of Epidemiology* **41**, 930–940 (2012).

218. Anderson, C. A. et al. Data quality control in genetic case-control association studies. *Nature Protocols* **5**, 1564-1573 (2010).

219. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* **81**, 559-575 (2007).

220. Chang, C.C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).

221. Das, S. et al. Next-generation genotype imputation service and methods. *Nature Genetics* **48**, 1284-1287 (2016).

222. McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nature Genetics* **48**, 1279-1283 (2016).

223. Loh, P. R. et al. Reference-based phasing using the Haplotype Reference Consortium panel. *Nature Genetics* **48**, 1443-1448 (2016).

224. Fuchsberger, C. et al. minimac2: faster genotype imputation. *Bioinformatics* **31**, 782-784 (2015).

225. 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**, 189–196 (2013).

226. Johnson, W. E. et al. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118-127 (2007).

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

228. Binder, A.M. (2021). *QC steps.* EpiCentered Research. https://www.epicenteredresearch.com/pace/qcsteps/.

229. McCarthey, D.L. et al. Identification of polymorphic and off-target probe binding sites on the Illumina Infinium MethylationEPIC BeadChip. *Genomics Data* **9**, 22-24 (2016).

230. Fortin, J. P. et al. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. *Bioinformatics* **33**, 558-560 (2017).

231. Triche, T. J. Jr. et al. Low-level processing of Illumina Infinium DNA Methylation BeadArrays. *Nucleic Acids Research* **41**, e90 (2013).

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

233. Houseman, E. A. et al. Reference-free deconvolution of DNA methylation data and mediation by cell composition effects. *BMC Bioinformatics* **17**, 259 (2016).

234. Venables, W. N. & Ripley, B. D. (2002). *Modern Applied Statistics with S.* Springer.

235. 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).

236. Newschaffer, C. J. et al. Infant siblings and the investigation of autism risk factors. *Journal of Neurodevelopmental Disorders* **4**, 7 (2012).

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

238. 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).

239. Aguilar-Lacasaña, S. et al. Methylenetetrahydrofolate Reductase (MTHFR) Gene Polymorphism and Infant's Anthropometry at Birth. *Nutrients* **13**, 831 (2021).

240. Michels, K. B. et al. Birthweight, maternal weight trajectories and global DNA methylation of LINE-1 repetitive elements. *PLOS One* **6**, e25254 (2011).

241. Dieckmann, L. et al. Characteristics of epigenetic aging across gestational and perinatal tissues. *Clinical Epigenetics* **13**, 97 (2021).

242. Hertz-Picciotto, I. et al. A Prospective Study of Environmental Exposures and Early Biomarkers in Autism Spectrum Disorder: Design, Protocols, and Preliminary Data from the MARBLES Study. *Environmental Health Perspectives* **126**, 117004 (2018).

243. Gilbert-diamond, D. et al. Relation between in Utero Arsenic Exposure and Birth Outcomes in a Cohort of Mothers and Their Newborns from New Hampshire. *Environmental Health Perspectives* **124**, 1299–1307 (2016).

244. Appleton, A. A. et al. Prenatal Programming of Infant Neurobehaviour in a Healthy Population. *Paediatric and Perinatal Epidemiology* **30**, 367–375 (2016).

245. Magi, R. & Morris, A. P. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics* **11**, 288 (2010).

246. Triche, Jr. T (2014). *FDb. InfiniumMethylation.hg19: Annotation package for Illumina Infinium DNA methylation probes. R package version 2.2.0.*

247. Hansen, K. D. (2016). *IlluminaHumanMethylation450kanno.ilmn12.hg19: Annotation for Illumina's 450k methylation arrays. R package version 0.6.0.*

248. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nature Methods* **9**, 215–216 (2012).

249. Roadmap Epigenomics Consortium et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–329 (2015).

250. Hamada, H. et al. Allele-Specific Methylome and Transcriptome Analysis Reveals Widespread Imprinting in the Human Placenta. *American Journal of Human Genetics*. **99**, 1045–1058 (2016).

251. Kamburov, A. et al. ConsensusPathDB: Toward a more complete picture of cell biology. *Nucleic Acids Research* **39**, 712–717 (2011).

252. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* **28**, 27-30 (2000).

253. Fabregat, A. et al. Reactome pathway analysis: a high-performance in-memory approach. *BMC Bioinformatics* **18**, 142 (2017).

254. Pico, A. R. et al. WikiPathways: Pathway Editing for the People. *PLOS Biol* (2008).

255. Lachmann, A. et al. ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. *Bioinformatics* **26**, 2438-44 (2010).

256. Chen, E. Y. et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14** (2013).

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

258. 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. *Human Molecular Genetics* **27**, 742–775 (2018).

259. 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. *Human Molecular Genetics* **24**, 1155–1168 (2015).

260. Lawrence, M. et al. Software for Computing and Annotating Genomic Ranges. *PLOS Computational Biology* **9**, e1003118 (2013).

261. Taylor-Weiner, A. et al. Scaling computational genomics to millions of individuals with GPUs. *Genome Biology* **20**, 228 (2019).

262. Ongen, H. et al. Fast and efficient QTL mapper for thousands of molecular phenotypes. *Bioinformatics* **32**, 1479-1485 (2016).

263. Min, J. L. et al. Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. *Nature Genetics* **53**, 1311-1321 (2021).

264. Gibbs, J. R. et al. Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLOS Genetics* **6**, e1000952 (2010).

265. Drong, A. W. et al. The presence of methylation quantitative trait loci indicates a direct genetic influence on the level of DNA methylation in adipose tissue. *PLOS ONE* **8**, e55923 (2013).

266. Olsson, A. H. et al. Genome-wide associations between genetic and epigenetic variation in-fluence mRNA expression and insulin secretion in human pancreatic islets. *PLOS Genetics* **10**, e1004735 (2014).

267. GTEx Consortium; Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis Working Group; Statistical Methods groups—Analysis Working Group; Genetic effects on gene expression across human tissues. *Nature* **550**, 204-213 (2017).

268. Davis, J. R. et al. An Efficient Multiple-Testing Adjustment for eQTL Studies that Accounts for Linkage Disequilibrium between Variants. *American Journal of Human Genetics* **98**, 216-224 (2016).

269. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 9440–9445 (2003).

270. Storey, J. D. et al. (2023). qvalue: Q-value estimation for false discovery rate control. doi:10.18129/B9.bioc.qvalue, R package version 2.34.0, https://bioconductor.org/packages/qvalue

271. Hansen, K. D. IlluminaHumanMethylationEPICanno.ilm10b4.hg19: Annotation for Illumina's EPIC methylation arrays. R package version 0.6.0, https://bitbucket.com/kasperdanielhansen/Illumina_EPIC (2017).

272. Breeze, C. E. et al. eFORGE: A Tool for Identifying Cell Type-Specific Signal in Epigenomic Data. *Cell Reports* **17**, 2137-2150 (2016).

273. Breeze, C. E. et al. eFORGE v2.0: updated analysis of cell type-specific signal in epigenomic data. *Bioinformatics* **35**, 4767-4769 (2019).

274. Breeze, C. E. Cell Type-Specific Signal Analysis in Epigenome-Wide Association Studies. *Methods in Molecular Biology* (Clifton, N.J.). **2432**, 57-71 (2022).

275. Maksimovic, J. et al. SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. *Genome Biology* **13**, R44 (2012).

276. Phipson, B. & Oshlack, A. DiffVar: a new method for detecting differential variability with application to methylation in cancer and aging. *Genome Biology* **15**, 465 (2014).

277. Maksimovic, J. et al. A. Removing unwanted variation in a differential methylation analysis of Illumina HumanMethylation450 array data. *Nucleic Acids Research* **43**, e106 (2015).

278. Phipson, B. et al. missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. *Bioinformatics* **32**, 286-288 (2016).

279. Yu, G. et al. DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. *Bioinformatics* **31**, 608-609 (2015).

280. Demontis, D. et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nature Genetics* **51**, 63-75 (2019).

281. Pappa, I. et al. A genome-wide approach to children's aggressive behavior: The EAGLE consortium. *American Journal Medicine Genetics Part B: Neuropsychiatric Genetics* **171**, 562-572 (2016).

282. Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium. Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. *Molecular Autism* **8**, 21 (2017).

283. Mullins, N. et al. Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nature Genetics* **53**, 817-829 (2021).

284. Jami, E. S. et al. Genome-wide Association Meta-analysis of Childhood and Adolescent Internalizing Symptoms. *Journal of the American Academy of Child and Adolescence Psychiatry* **61**, 934-945 (2022).

285. Howard, D. M. et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nature Neuroscience* **22**, 343-352 (2019).

286. International Obsessive Compulsive Disorder Foundation Genetics Collaborative (IOCDF-GC) and OCD Collaborative Genetics Association Studies (OCGAS). Revealing the complex genetic architecture of obsessive-compulsive disorder using meta-analysis. *Molecular Psychiatry* **23**, 1181-1188 (2018).

287. Forstner, A. J. et al. Genome-wide association study of panic disorder reveals genetic overlap with neuroticism and depression. *Molecular Psychiatry* **26**, 4179-4190 (2021).

288. Jiang, L. et al. A generalized linear mixed model association tool for biobank-scale data. *Nature Genetics* **53**, 1616-1621 (2021).

289. Trubetskoy, V. et al. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**, 502-508 (2022).

290. Boughton, A. P. et al. LocusZoom.js: interactive and embeddable visualization of genetic association study results. *Bioinformatics* **37**, 3017-3018 (2021).

291. Wu, Y. et al. Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. *Nature Communications* **9**, 918 (2018).

292. Yang, J. et al. GCTA: a tool for genome-wide complex trait analysis. *American Journal of Human Genetics* **88**, 76–82 (2011).

293. Yang, J. et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nature Genetics* **44**, 369–375 (2012).

294. Love, M. I. et al. Modeling of RNA-seq fragment sequence bias reduces systematic errors in transcript abundance estimation. *Nature Biotechnology* **34**, 1287–1291 (2016).

295. Peng, S. et al. Expression quantitative trait loci (eQTLs) in human placentas suggest developmental origins of complex diseases. *Human Molecular Genetics* **26**, 3432–3441 (2017).

296. Gillespie, M. et al. The reactome pathway knowledgebase 2022. *Nucleic Acids Research* **50**, D687-D692 (2022).

297. Schriml, L. M. et al. Human Disease Ontology 2018 update: classification, content and workflow expansion. *Nucleic Acids Research* **47**, D955-D962 (2019).

298. Garrido-Martín, D. et al. A fast non-parametric test of association for multiple traits. *Genome Biology* **24**, 230 (2023).

299. Psychiatric Genomics Consortium. (2023, January 23). *PGC. Psychiatric Genomics Consortium*. https://pgc.unc.edu.

300. Beate St Pourcain. (2023, January 23). *EAGLE Consortium. EArly Genetics and Lifecourse Epidemiology.* Max Planck Institute for Psycholinguistics, Language and Genetics Department. https://www.eagle-consortium.org.

301. Lee, Y. et al. Placental epigenetic clocks: estimating gestational age using placental DNA methylation levels. *Aging (Albany NY)* **11**, 4238-4253 (2019).

302. Shabalin, A. A. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* **28**, 1353-1358 (2012).

303. Johnston, K. J. A. et al. Identification of novel common variants associated with chronic pain using conditional false discovery rate analysis with major depressive disorder and assessment of pleiotropic effects of LRFN5. *Translational Psychiatry* **9**, 310 (2019).

304. Mischler, A. et al. Two distinct trophectoderm lineage stem cells from human pluripotent stem cells. *Journal of Biological Chemistry* **296**, 100386 (2021).

305. Tang, M. et al. circRNA-DURSA regulates trophoblast apoptosis via miR-760-HIST1H2BE axis in unexplained recurrent spontaneous abortion. *Molecular Therapy Nucleic Acids* **26**, 1433-1445 (2021).

306. Liu, Y. et al. Alcohol exposure alters DNA methylation profiles in mouse embryos at early neurulation. *Epigenetics* **4**, 500-511 (2009).

307. Kent, W. J. et al. The human genome browser at UCSC. *Genome Research* **12**, 996-1006 (2002).

308. Cleal, J. K. et al. A systems perspective on placental amino acid transport. *Journal of Physiology* **596**, 5511-5522 (2018).

309. Mercer, G. V. et al. Identifying placental metabolic biomarkers of preterm birth using nuclear magnetic resonance of intact tissue samples. *Placenta* **143**, 80–86 (2023).

310. Brkić, J. et al. Differential Role of Smad2 and Smad3 in the Acquisition of an Endovascular Trophoblast-Like Phenotype and Preeclampsia. *Frontiers in Endocrinology (Lausanne)* **11**, 436 (2020).

311. Liang, M. H. et al. Lithium inhibits Smad3/4 transactivation via increased CREB activity induced by enhanced PKA and AKT signaling. *Molecular and Cellular Neuroscience* **37**, 440-453 (2008).

312. Qi, T. et al. Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. *Nature Communications*. **9**, 2282 (2018).

313. Jaffe, A. E. et al. Mapping DNA methylation across development, genotype and schizophrenia in the human frontal cortex. *Nature Neuroscience* **19**, 40-47 (2016).

314. Decker, M. et al. EH3 (ABHD9): the first member of a new epoxide hydrolase family with high activity for fatty acid epoxides. *Journal of Lipid Research* **53**, 2038–2045 (2012)

315. De Taeye, B. M. et al. Expression and regulation of soluble epoxide hydrolase in adipose tissue. *Obesity* **18**, 489–498 (2010).

316. Sun, L. et al. CRISPR/Cas9 mediated establishment of a human CSRP3 compound heterozygous knockout hESC line to model cardiomyopathy and heart failure. *Stem Cell Research* **49**, 102077 (2020).

317. Hall, E. et al. The effects of high glucose exposure on global gene expression and DNA methylation in human pancreatic islets. *Molecular and Cellular and Endocrinology* **472**, 57–67 (2018).

318. Usha Kalyani, R. et al. Fer1L5, a Dysferlin Homologue Present in Vesicles and Involved in C2C12 Myoblast Fusion and Membrane Repair. *Biology* **9**, 386 (2020).

319. Knöfler, M. et al. Human placenta and trophoblast development: key molecular mechanisms and model systems. *Cellular and Molecular Life Sciences* **76**, 3479–3496 (2019).

320. Almomani, S. N. et al. Identification and validation of DNA methylation changes in pre-eclampsia. *Placenta* **110**, 16–23 (2021).

321. Diceglie, C. et al. Placental Antioxidant Defenses and Autophagy-Related Genes in Maternal Obesity and Gestational Diabetes Mellitus. *Nutrients* **13**, 1303 (2021).

322. Zou, Y. et al. Association of a gain-of-function variant in LGR4 with central obesity. *Obesity (Silver Spring)* **25**, 252–260 (2017).

323. Coorens, T. H. H. et al. Inherent mosaicism and extensive mutation of human placentas. *Nature* **592**, 80–85 (2021).

324. Costanzo, V. et al. Exploring the links between cancer and placenta development. *Open Biology* **8**, 180081 (2018).

325. Calabuig-Navarro, V. et al. Effect of Maternal Obesity on Placental Lipid Metabolism. *Endocrinology* **158**, 2543–2555 (2017).

326. Alcala, M. et al. Antioxidants and Oxidative Stress: Focus in Obese Pregnancies. *Frontiers in Physiology* **9**, 1569 (2018).

327. Malti, N. et al. Oxidative stress and maternal obesity: feto-placental unit interaction. *Placenta* **35**, 411–416 (2014).

328. Yuen, R. K. et al. Hypoxia alters the epigenetic profile in cultured human placental trophoblasts. *Epigenetics* **8**, 192–202 (2013).

329. Fernandez-Twinn, D. S. et al. Exercise rescues obese mothers' insulin sensitivity, placental hypoxia and male offspring insulin sensitivity. *Scientific Reports* **7**, 44650 (2017).

330. Fernandez-Jimenez, N. et al. Comparison of Illumina 450 K and EPIC arrays in placental DNA methylation. *Epigenetics* **14**, 1177–1182 (2019).

331. Casas, M. et al. Maternal pre-pregnancy overweight and obesity, and child neuropsychological development: two Southern European birth cohort studies. *International Journal of Epidemiology* **42**, 506–517 (2013).

332. Chen, J. et al. Fast and robust adjustment of cell mixtures in epigenome-wide association studies with Smart SVA. *BMC Genomics* **18**, 413 (2017).

333. Rao, S. et al. Correction: Analysis of genetic differences between psychiatric disorders: exploring pathways and cell types/tissues involved and ability to differentiate the disorders by polygenic scores. *Translational Psychiatry* **12**, 237 (2022).

334. Romero, C. et al. Exploring the genetic overlap between twelve psychiatric disorders. *Nature Genetics* **54**, 1795-1802 (2022).

335. Grotzinger, A. D. et al. Genetic architecture of 11 major psychiatric disorders at biobehavioral, functional genomic and molecular genetic levels of analysis. *Nature Genetics* **54**, 548-559 (2022).

336. Bora, E. Differences in cognitive impairment between schizophrenia and bipolar disorder: Considering the role of heterogeneity. *Psychiatry and Clinical Neurosciences* **70**, 424-433 (2016).

337. Esteki-Zadeh., A. et al. Human cytomegalovirus infection is sensitive to the host cell DNA methylation state and alters global DNA methylation capacity. *Epigenetics* **7**, 585-593 (2012).

338. Elgueta, D. et al. Consequences of Viral Infection and Cytokine Production During Pregnancy on Brain Development in Offspring. *Frontiers in Immunology* **13**, 816619 (2022).

339. Ursini, G. et al. Prioritization of potential causative genes for schizophrenia in placenta. *Nature Communications* **14**, 2613 (2023).

340. Han, V. X. et al. Maternal immune activation and neuroinflammation in human neurodevelopmental disorders. *Nature Reviews Neurology* **17**, 564-579 (2021).

341. Weber-Stadlbauer, U. Epigenetic and transgenerational mechanisms in infection-mediated neurodevelopmental disorders. *Translational Psychiatry* **7**, e1113 (2017).

342. Richetto, J. et al. Genome- wide DNA methylation changes in a mouse model of infection- mediated neurodevelopmental disorders. *Biological Psychiatry* **81**, 265–276 (2017).

343. Knuesel, I. et al. Maternal immune activation and abnormal brain development across CNS disorders. *Nature Reviews Neurology* **10**, 643–660 (2014).

344. Banik, A. et al. Maternal factors that induce epigenetic changes contribute to neurological disorders in offspring. *Genes* **8**, 150 (2017).

345. Hodyl, N. A. et al. Cord blood DNA methylation biomarkers for predicting neurodevelopmental outcomes. *Genes* **7**, 117 (2016).

346. Howerton, C. L. & Bale, T. L. Prenatal programing: at the intersection of maternal stress and immune activation. *Hormones and Behavior* **62**, 237–242 (2012).

347. Apicella, C. et al. Pan-Genomic Regulation of Gene Expression in Normal and Pathological Human Placentas. *Cells* **12**, 578 (2023)

348. Zhang, J. et al. SFMBT1 functions with LSD1 to regulate expression of canonical histone genes and chromatin-related factors. *Genes & Development* **27**, 749-66 (2013).

349. Miri, K. et al. The imprinted polycomb group gene Sfmbt2 is required for trophoblast maintenance and placenta development. *Development* **140**, 4480-4489 (2013).

350. Hannon, E. et al. Elevated polygenic burden for autism is associated with differential DNA methylation at birth. *Genome Medicine* **10**, 19 (2018).

351. Walton, E. et al. Epigenetic profiling of ADHD symptoms trajectories: a prospective, methylome-wide study. *Molecular Psychiatry* **22**, 250–256 (2017).

352. Lybaek, H. et al. LRFN5 locus structure is associated with autism and influenced by the sex of the individual and locus conversions. *Autism Research* **15**, 421-433 (2022).