

## Potentially causal associations between placental DNA methylation and schizophrenia and other neuropsychiatric disorders

Ariadna Cilleros-Portet<sup>1</sup>, Corina Lasseur<sup>2</sup>, Sergi Marí<sup>1</sup>, Marta Cosin-Tomas<sup>3,4,5</sup>, Manuel Lozano<sup>6,7</sup>, Amaia Irizar<sup>4,8,9</sup>, Amber Burt<sup>10</sup>, Iraia García-Santisteban<sup>1</sup>, Diego Garrido Martín<sup>11,12</sup>, Geòrgia Escaramís<sup>4,13</sup>, Alba Hernangomez-Laderas<sup>1</sup>, Raquel Soler-Blasco<sup>4,6,14</sup>, Charles E. Breeze<sup>15</sup>, Bárbara P. Gonzalez-Garcia<sup>1</sup>, Loreto Santa-Marina<sup>4,9,16</sup>, Jia Chen<sup>2</sup>, Sabrina Llop<sup>4,6</sup>, Mariana F. Fernández<sup>4,17</sup>, Martine Vrijhed<sup>3,4,5</sup>, Jesús Ibarluzea<sup>4,9,16</sup>, Mònica Guxens<sup>3,4,5,18</sup>, Carmen Marsit<sup>10</sup>, Mariona Bustamante<sup>3,4,5</sup>, Jose Ramon Bilbao<sup>1,19</sup>, Nora Fernandez-Jimenez<sup>1\*</sup>

<sup>1</sup>*Department of Genetics, Physical Anthropology and Animal Physiology, Biocruces-Bizkaia Health Research Institute and University of the Basque Country (UPV/EHU), Leioa, Spain.*

<sup>2</sup>*Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, NY, USA.*

<sup>3</sup>*ISGlobal, Barcelona, Spain.*

<sup>4</sup>*Spanish Consortium for Research on Epidemiology and Public Health (CIBERESP), Instituto de Salud Carlos III, 28029, Madrid, Spain.*

<sup>5</sup>*Universitat Pompeu Fabra, Barcelona, Spain.*

<sup>6</sup>*Epidemiology and Environmental Health Joint Research Unit, FISABIO-Universitat Jaume I-Universitat de València, Valencia, Spain.*

<sup>7</sup>*Preventive Medicine and Public Health, Food Sciences, Toxicology and Forensic Medicine Department, Universitat de València, Valencia, Spain.*

<sup>8</sup>*Department of Preventive Medicine and Public Health, University of the Basque Country (UPV/EHU), Leioa, Spain.*

<sup>9</sup>*Biodonostia Health Research Institute, 20013, San Sebastian, Spain.*

<sup>10</sup>*Gangarosa Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, GA, USA.*

<sup>11</sup>*Department of Genetics, Microbiology and Statistics, Faculty of Biology, Universitat de Barcelona (UB), 08028 Barcelona, Spain.*

<sup>12</sup>*Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, 08003 Barcelona, Spain.*

<sup>13</sup>*Departament de Biomedicina, Facultat de Medicina i Ciències de la Salut, Institut de Neurociències, Universitat de Barcelona, Casanova 143, Barcelona, Spain.*

<sup>14</sup>*Department of Nursing, Universitat de València, Valencia, Spain.*

<sup>15</sup>*UCL Cancer Institute, University College London, 72 Huntley St, London WC1E 6DD, United Kingdom.*

<sup>16</sup>*Department of Health of the Basque Government, Subdirectorate of Public Health of Gipuzkoa, Avenida Navarra 4, 20013, San Sebastian, Spain.*

<sup>17</sup>*Biomedical Research Center (CIBM) & Department of Radiology and Physical Medicine, School of Medicine University of Granada, 18016 Granada, Spain; Instituto de Investigación Biosanitaria de Granada (ibs.GRANADA), 18012 Granada, Spain.*

<sup>18</sup>*Department of Child and Adolescent Psychiatry/Psychology, Erasmus MC, University Medical Centre, Rotterdam, The Netherlands.*

<sup>19</sup>*CIBER de Diabetes y Enfermedades Metabólicas Asociadas (CIBERDEM), Madrid, Spain.*

\*Corresponding author:

Nora Fernandez-Jimenez, PhD

Assistant Professor in Medical Genetics

Immunogenetics Research Laboratory (IRLab)

Department of Genetics, Physical Anthropology and Animal Physiology

María Goyri building - lab 2.06

Faculty of Medicine and Nursing - office AM7.2

Biocruces-Bizkaia Health Research Institute

University of the Basque Country - UPV/EHU

[nora.fernandez@ehu.eus](mailto:nora.fernandez@ehu.eus)

tel.: (+34) 94 601 7765 (lab) / 2909 (office)

## **Abstract**

Increasing evidence supports the role of placenta in neurodevelopment and potentially, in the later onset of neuropsychiatric disorders. Recently, methylation quantitative trait loci (mQTL) and interaction QTL (iQTL) maps have proven useful to understand SNP-genome wide association study (GWAS) relationships, otherwise missed by conventional expression QTLs. In this context, we propose that part of the genetic predisposition to complex neuropsychiatric disorders acts through placental DNA methylation (DNAm). We constructed the first public placental *cis*-mQTL database including nearly eight million mQTLs calculated in 368 fetal placenta DNA samples from the INMA project, ran cell type- and gestational age-imQTL models and combined those data with the summary statistics of the largest GWAS on 10 neuropsychiatric disorders using Summary-based Mendelian Randomization (SMR) and colocalization. Finally, we evaluated the influence of the DNAm sites identified on placental gene expression in the RICHS cohort. We found that placental *cis*-mQTLs are highly enriched in placenta-specific active chromatin regions, and useful to map the etiology of neuropsychiatric disorders at prenatal stages. Specifically, part of the genetic burden for schizophrenia, bipolar disorder and major depressive disorder confers risk through placental DNAm. The potential causality of several of the observed associations is reinforced by secondary association signals identified in conditional analyses, regional pleiotropic methylation signals associated to the same disorder, and cell type-imQTLs, additionally associated to the expression levels of relevant immune genes in placenta. In conclusion, the genetic risk of several neuropsychiatric disorders could operate, at least in part, through DNAm and associated gene expression in placenta.

## **Introduction**

The impact of the intrauterine environment on development and health, both fetal and long-term, has been known for many years<sup>1</sup>. In this line, the developmental origins of health and disease (DOHaD) hypothesis postulates that perinatal and early life environments can impact fetal and later-life health<sup>2</sup>. Prenatal stress affects the quality of the intrauterine environment, and is highly related to cardiovascular and metabolic, as well as behavioral and neurodevelopmental disorders<sup>3</sup>. Thus, the placenta is an ephemeral fetal organ that is uniquely situated to evaluate prenatal exposures in the context of DOHaD, because it manages the transport of nutrients, oxygen, waste, and endocrine signals between mother and fetus<sup>4</sup>. Additionally, the placenta constitutes the interface between mother and child during pregnancy and is the maximum regulator of the prenatal milieu, having a key role in the growth and the neurodevelopment of the fetus<sup>5</sup>. Hence, the placenta has been described as the third brain that links the fetal brain with the mature maternal brain, thus becoming the cornerstone to understand the prenatal environmental effects on neurodevelopment<sup>6</sup>, and potentially, on the appearance of neurodevelopmental and neuropsychiatric disorders later in life.

Particularly in the case of schizophrenia (SCZ), many susceptibility genes have been identified, demonstrating a remarkable genetic basis, but a considerable amount of research suggest that environmental factors may also play a role<sup>7</sup>. The neurodevelopmental hypothesis of SCZ was first proposed by Daniel Weinberger in 1987 and has been reinforced in the last decades, with increasing evidence of neurodevelopmental abnormalities contributing to the pathophysiology of the disease<sup>8,9,10</sup>. The central argument of this hypothesis states that abnormal fetal neurodevelopment creates a vulnerability to develop SCZ later in life. In fact, experimental evidence states that prenatal insults such as maternal immune activation (MIA) are associated with SCZ in offspring<sup>11</sup>. MIA occurs when inflammatory markers rise above the normal range in pregnancy as a result of maternal inflammation and can be caused by psychosocial stress, infection, or other factors<sup>12</sup>.

Besides, there are many studies proposing that the intrauterine environment alters the placental function through epigenetic mechanisms, such as DNA methylation (DNAm)<sup>13</sup>. It is very important to note that, while DNAm is bimodally distributed in most tissues and cell types, it follows a trimodal distribution in the very specific case of placenta, due to its high content of both partially methylated domains and CpG positions with intermediate methylation levels<sup>14</sup>. Although a very recent study has pointed out that DNAm of healthy donors could be highly cell type-specific, with modest impact of either genetics or other factors<sup>15</sup>, it is worth to mention that DNAm has been considered as a bridge between the environment and the genome that at the same time, is under the control of both environmental and genetic factors. The peculiarity of placenta as an ephemeral organ that connects two organisms, together with its very particular methylome that is remarkably enriched in CpG sites with intermediate methylation levels<sup>14</sup>, leads us to speculate that the genome-environment interaction and its impact on DNAm is very likely unique in placenta and deserves further investigation.

In the past several years, different works have demonstrated that placental DNAm is sensitive to environmental factors surrounding the gestation. In 2021 the Pregnancy and Childhood Epigenetics (PACE) consortium conducted a meta-analysis on 1,700 placental samples and found a placenta-specific DNAm signature of maternal smoking during pregnancy, with differentially methylated CpG sites located in active regions of the placental epigenome, and close to genes involved in the regulation of inflammatory activity, growth factor signaling, and response to environmental stressors<sup>16</sup>. One year later, the same consortium reported in the largest placental DNAm study conducted to date that maternal pre-pregnancy body mass index also impacts the placental methylome, specifically at CpG sites located close to obesity-related genes and altogether, enriched in oxidative-stress and cancer pathways<sup>17</sup>.

Noting the impact of the genetic background on placental DNAm, several studies have reported a number of methylation quantitative trait loci (mQTL) in placenta and have used them in different downstream applications<sup>18,19,20,21</sup>. Among others, Tekola-Ayele and collaborators showed candidate functional pathways that underpin the genetic regulation of birth weight via placental epigenetic and transcriptomic mechanisms<sup>20</sup>. However, the catalogue of mQTLs they provided was limited to birth weight-related genomic loci. More recently, Casazza reported nearly 50,000 placental *cis*-mQTLs and 2,489 sex-specific placental *cis*-mQTLs<sup>21</sup>. They found out that placental mQTLs were enriched in genome-wide association study (GWAS) loci for both growth- and immune-related traits, and that male- and female-specific mQTLs were more abundant than non-sex-specific ones. Another remarkable discovery is that they found a modest enrichment of placental mQTLs in proportion to the calculated SNP heritability in the case of neuropsychiatric disorders compared to immune- and growth-related traits. However, they did not rule out the possibility that a part of the genetic susceptibility of neuropsychiatric disorders is conferred through modification of the placenta, and if this was the case, it would be relevant not only in terms of etiopathology, but also from a clinical viewpoint, due to the importance of localizing therapeutic targets in the right tissues, contexts and developmental stages.

In cross-tissue analyses, mQTLs have been highlighted as powerful instruments revealing molecular links to traits otherwise missed by expression quantitative trait loci (eQTL). In fact, recent work by Oliva et al. reported that trait-associated variants are more likely to result in detectable changes in DNAm rather than gene expression, highlighting the relevance of multi-tissue mQTL maps<sup>22</sup>. On the other hand, in 2021, cell type-interacting eQTLs were defined as proxies of cell type-specific eQTLs and were found not to be covered by standard eQTLs, while they appeared to be highly valuable for gaining a mechanistic understanding of complex trait associations<sup>23</sup>. Taken together, placenta-specific mQTLs as well as placenta cell type interacting mQTLs (imQTLs) can be useful tools in the search for the etiology of genetic associations. In this sense, it is of utmost importance to make this type of resources public and available to the scientific community.

Taking into consideration i) the increasing evidence supporting the role of placenta in neurodevelopment and potentially, in the onset of neuropsychiatric traits and disorders, ii) the peculiarity of the placental methylome, and iii) the potential of multi-tissue mQTL and imQTL maps to clarify the etiology of complex traits, we proposed that part of the genetic predisposition to complex neuropsychiatric disorders acts through the placental methylome. Thus, we constructed the first publicly available placental *cis*-mQTL database including nearly 8 million nominal mQTLs calculated in 368 fetal placenta samples, investigated cell type- and gestational age (GA)-imQTLs and integrated all these data with summary statistics of the largest GWAS on 10 neuropsychiatric disorders, using Summary-based Mendelian Randomization (SMR) and colocalization approaches. Finally, we evaluated the functional role of the identified DNAm sites on placental gene expression. We found that placental *cis*-mQTLs are enriched in placental active genomic regions, and are useful to map the etiology of neuropsychiatric disorders to prenatal stages. Specifically, part of the SCZ, bipolar disorder (BIP) and major depressive disorder (MDD) genetic risk could act through placental DNAm. The potential causality of several associations observed is reinforced by secondary association signals identified in conditional analyses, regional pleiotropic methylation signals associated to the same disorder, as well as cell type imQTLs involved, that additionally associate to the expression levels of relevant genes in placenta.

## Results

### Placental *cis*-mQTL characterization

The nominal mQTL database was calculated in 368 fetal placenta DNA samples from the Gipuzkoa, Sabadell and València cohorts of the *Infancia y Medio Ambiente* (INMA) project<sup>24</sup>, and contained 7,921,914 *cis*-mQTLs ( $P_{\text{nominal}} < 5 \times 10^{-8}$ ) with 110,721 and 1,900,580 unique CpGs and SNPs, respectively. Briefly, placental DNAm was modeled as a function of genotype in 0.5 Mb windows, with fetal sex, 5 genotype principal components (PC) and Planet-estimated cell types<sup>25</sup> as covariates. Phenotype data of the donor mothers is summarized in Table 1. The vast majority of the SNP-CpG pairs in *cis*-mQTLs were located close to each other with a median distance of 44 kb, indicating that genetically modulated DNAm is typically close to the implicated regulatory variant, as previously described by other authors<sup>26</sup> (Figure 1A). The distribution of mQTLs across chromosomes was in line with chromosome length, except for chromosome 6, where a higher number of CpG-SNP pairs was observed (883,548 CpG-SNP pairs with 8,173 and 153,743 unique CpGs and SNPs, respectively) (Figure 1B). This was expected due to the complex linkage disequilibrium (LD) structure of the Human Leucocyte Antigen (HLA) region on this chromosome. Additionally, placental *cis*-mQTLs presented a uniform distribution of effect sizes and directions (Figure 1C).

In general, mQTL-CpGs were depleted from CpG islands and promoters ( $P < 2.2 \times 10^{-16}$  and  $P < 2.2 \times 10^{-16}$ , respectively), and enriched within gene bodies ( $P = 1.71 \times 10^{-8}$ ) and genomic features showing intermediate methylation values such as open sea, and CpG island shelf and shore regions ( $P < 2.2 \times 10^{-16}$ ,  $P = 3.83 \times 10^{-7}$  and  $P < 2.2 \times 10^{-16}$ , respectively) (Figure 1D, E and F, and Supplementary Data 1). Using eFORGE<sup>27,28,29</sup>, we were able to detect an enrichment in fetal placenta-specific DNase I hotspots (Figure 1G) and H3K4me1 broadPeaks (Supplementary Figure 1A and Supplementary Data 2), that mark accessible chromatin and enhancer regions, respectively. Afterwards, we annotated the mQTL-CpGs to their closest genes, performed a gene set enrichment analysis with the Disease Ontology database<sup>30</sup>, and obtained 140 enriched gene sets, including developmental disorder of mental health (Benjamini-Hochberg  $P = 1.41 \times 10^{-10}$ ), psychotic disorder (Benjamini-Hochberg  $P = 1.41 \times 10^{-10}$ ) and SCZ (Benjamini-Hochberg  $P = 1.41 \times 10^{-10}$ ) (Supplementary Data 3 and Supplementary Figure 2).

Apart from the nominal *cis*-mQTLs, permuted and conditional *cis*-mQTLs were also calculated in our 368 placenta DNA samples. Both permuted and conditional mQTL databases showed a median SNP-CpG distance of 7kb, remarkably shorter than in the nominal database. Chromosome 6 was no longer the one showing the highest absolute number of *cis*-mQTLs (Supplementary Figures 3 and 4). These two observations were expected considering that both permuted and conditional mQTL analyses give as a result a reduced number of SNPs per CpG compared to the nominal approach. Additionally, we observed enrichments and depletions similar to those reported for the nominal database (Supplementary Figures 1, 3 and 4, and Supplementary Data 1, 2 and 3). When we compared the *cis*-mQTLs included in the three different databases, the vast majority of the mQTLs were common. The nominal database computes all the possible mQTLs in the SNP-CpG window. Therefore, as expected, it contained a much larger number of mQTLs than the two other databases (Supplementary Figure 5).

The three complete placental *cis*-mQTL databases are publicly available online in the following address: <https://smari.shinyapps.io/shi>

### Placental cell type- and GA- imQTLs

The most abundant cell type in fetal placenta are syncytiotrophoblasts (STB). During gestation, undifferentiated trophoblasts (TB) change into fully differentiated STB, a continuous, specialized layer of epithelial cells<sup>31</sup>. We estimated the cell-type proportions of our samples by using the reference-based

method Planet<sup>25</sup>. As expected, the estimated content of STB was negatively correlated with the estimated proportion of TB in our placental samples ( $P < 2.2 \times 10^{-16}$ ,  $R^2 = -0.89$ ). In turn, STB content was positively correlated with gestational age (GA) ( $P < 4.5 \times 10^{-5}$ ,  $R^2 = 0.21$ ) (Figure 2A and 2B). We calculated STB-, TB- and GA-imQTLs, as proxies for STB-, TB- and GA-specific *cis*-mQTLs, and obtained 202 positive and 728 negative STB-imQTLs, 237 positive and 594 negative TB-imQTLs, and 412 GA-imQTLs ( $P_{\text{interaction}} < 5 \times 10^{-8}$ ) (Supplementary Data 4). The higher amount of STB-imQTLs, revealed a higher statistical power for the most abundant cell type, as previously stated by Kim-Hellmuth et al<sup>23</sup>. Out of the 930 STB-imQTLs and 831 TB-imQTLs, 443 were common to both cell types (Figure 2C). Considering the same allele for both STB and TB, the effect sizes of the shared cell type-imQTLs were negatively correlated ( $P < 2.2 \times 10^{-16}$ ,  $R^2 = -0.98$ ) with opposite effect directions as in the example in Figure 2D, E and F. Of note, most of the overlapping imQTLs were positively correlated with the STB content, and negatively correlated with the TB abundance (431/443). In contrast, no overlap was observed with GA-imQTLs (Figure 2C). The effect of GA on the genetic regulation of placental DNAm seemed to be independent of the cell type composition of each sample and therefore, rather a matter of time of gestation regardless of the TB to STB transition.

### **Multi-omics approaches to unravel the placental origin of neuropsychiatric traits and disorders**

We selected the largest GWAS for each of the traits analyzed, namely attention deficit and hyperactivity disorder (ADHD)<sup>32</sup>, aggressive behavior in children (AGR)<sup>33</sup>, autism spectrum disorder (ASD)<sup>34</sup>, BIP<sup>35</sup>, internalizing problems (INT)<sup>36</sup>, MDD<sup>37</sup>, obsessive-compulsive disorder (OCD)<sup>38</sup>, panic disorder (PD)<sup>39</sup>, suicidal attempt (SA)<sup>40</sup> and SCZ<sup>41</sup> (Table 2). Most of the studies were carried out in the context of either the Psychiatric Genomics Consortium (PGC)<sup>42</sup> or the Early Genetics and Lifecourse Epidemiology (EAGLE)<sup>43</sup> consortium. GWAS sample sizes ranged from 6,183 in the case of SA and 500,199 samples in MDD (Table 2).

We searched for placenta DNAm sites pleiotropically associated with neuropsychiatric disorders using the SMR and the heterogeneity in dependent instruments (HEIDI) tests included in the SMR software<sup>44</sup>. Briefly, SMR was originally developed to test for pleiotropic association between the expression level of a gene (or the methylation level of a CpG site) and a complex trait of interest using summary-level data from GWAS and QTL studies. We used the summary statistics of the aforementioned GWAS and the nominal placental *cis*-mQTL database. No significant hits were found for most traits, including early onset disorders such as ASD and ADHD, and others such as OCD or SA. In turn, multiple SMR hits (Bonferroni  $P_{\text{SMR}} < 0.05$  and  $P_{\text{HEIDI}} > 0.05$ ) were identified for BIP (n=29), MDD (n=27) and especially for SCZ (n=188) (Supplementary Data 5).

To replicate pleiotropic associations from the SMR analyses, we performed a colocalization test with the same data. The Bayesian colocalization analysis implemented in the *coloc* R package<sup>45</sup> focuses on finding the intersection between significant variants independently associated with two phenotypes, comparing the association patterns in the GWAS and QTL analyses across genomic regions, and thus combining the summary statistics into posterior probabilities for five hypotheses (see Methods for more information). Colocalization was performed across 64, 102, and 287 genomic regions defined in each GWAS associated with BIP, MDD and SCZ, respectively. These regions spanned 110,721 placental DNAm sites from the nominal *cis*-mQTL database. In BIP, the posterior probabilities for 47 regions, involving 488 DNAm sites in 537 region-CpG pairs, were supportive of a colocalization signal ( $\text{PPA4} > 0.8$ ) (Supplementary Data 6). In MDD, the posterior probabilities for 52 regions, involving 284 DNAm sites in 295 region-CpG pairs, were supportive of a colocalization signal ( $\text{PPA4} > 0.8$ ) (Supplementary Data 6). Finally, in SCZ, the posterior probabilities for 188 regions, involving 2,057 DNAm sites in 2,177 region-CpG pairs, were supportive of a colocalization signal ( $\text{PPA4} > 0.8$ ) (Supplementary Data 6). When the overlap with SMR hits was assessed, out of the 29 SMR hits in BIP,

22 DNAm sites (75%) showed evidence of colocalization with BIP. In MDD, from the 27 SMR hits, 13 DNAm sites (48%) showed evidence of colocalization with MDD. And, from the 188 SMR hits in SCZ, 96 DNAm sites (51%) showed evidence of colocalization with SCZ.

To provide more insights into the functional relevance of our findings, we examined associations between placental DNAm and placental gene expression for the CpGs from the SMR and colocalization analyses. We interrogated expression quantitative trait methylation sites (eQTM) from an independent set of 195 fetal placenta samples from the Rhode Island Child Health (RICHS) study<sup>46</sup>. Briefly, eQTM were calculated using linear models in MatrixEQTL, adjusted by fetal sex, 5 gene expression PCs and Planet-estimated cell types, and considering all the genes in a 0.5 Mb window up and downstream of each CpG. Among the 29, 27 and 188 SMR-significant DNAm sites identified in BIP, MDD and SCZ, we found 2, 4 and 48 significant eQTM-genes (FDR<0.05), respectively (Supplementary Data 7), strongly supporting the placental function of the CpGs identified.

Next, we identified the intersect between SMR, colocalization and eQTM results for each trait (Supplementary Figure 6). In the case of BIP, two placental DNAm sites that colocalized with two GWAS loci on chromosomes 3 and 5 showed DNAm levels correlated with the expression of two nearby genes (*ITIH4* and *ZNF592*). For MDD, two placental DNAm sites colocalized with a single MDD-associated genomic locus on chromosome 14, with DNAm levels correlating with the placental expression of *LRFN5*. Finally, in SCZ, 28 placental DNAm sites colocalized with 11 SCZ-associated loci in chromosomes 3, 6, 7, 11, 12, 18, 19 and 22, with DNAm levels that correlated with the placental expression of 19 nearby genes (*FXR1*, *GLB1L3*, *HLA-H*, *IRF3*, *NAGA*, *PGBD1*, *PSMG3*, *RNF39*, *SLC6A16*, *TOB2P1*, *TRIM27*, *TXNL4A*, *VARS2*, *VPS37B*, *ZKSCAN4*, *ZNF165*, *ZNRD1-AS1*, *ZSCAN12P1* and *ZSCAN23*).

One of the overlapping hits in BIP was cg17890764 (Bonferroni  $P_{SMR} = 0.041$ ) that colocalized with the BIP-associated region chr3:52214443-53175443 (PPA4 = 0.828). The DNAm levels of this CpG were negatively correlated with the placental expression of *ITIH4* (FDR  $P = 6.28 \times 10^{-4}$ ,  $R^2 = -0.302$ ), a gene that has been highlighted in an integrative analysis of BIP GWAS and different regulatory SNP annotation datasets, and is known to participate in neuroinflammation<sup>47</sup> (Supplementary Figure 7).

Regarding MDD, two independent, pleiotropically associated CpGs, cg23217097 and cg10318063 (Bonferroni  $P_{SMR} = 0.047$  and  $P_{SMR} = 0.004$ , respectively), colocalized with the MDD-associated locus on chr14:41940872-42476274 (PPA4 = 0.977 and PPA4 = 0.974, respectively) (Figure 4). The DNAm levels of those CpGs correlated with the expression of *LRFN5* in placenta (FDR  $P = 6.08 \times 10^{-5}$  and  $R^2 = -0.342$ , and FDR  $P = 1.03 \times 10^{-13}$  and  $R^2 = 0.538$ , respectively). 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<sup>48</sup>. 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<sup>49</sup>. Finally, the two independent CpGs identified are located in the promoter region of *LRFN5*. Altogether, these results suggest that the reported association is potentially causal, rather than pleiotropic.

In SCZ, cg03172226 was found to be pleiotropically associated with the disorder (Bonferroni  $P_{SMR} = 0.032$ ), colocalized with the GWAS peak on chr19:50007909-50252109 (PPA4 = 0.999), and DNAm levels of the CpG were correlated with placental expression of *IRF3* (FDR  $P = 0.018$  and  $R^2 = 0.237$ ) (Supplementary Figure 8). Importantly, *IRF3* is a regulator of type I interferons with a pivotal role in MIA<sup>50</sup>. Another example of pleiotropic association with SCZ is cg17805547 (Bonferroni  $P_{SMR} = 0.032$ ). This CpG colocalized with the SCZ-associated regions chr3:179745069-180418069 and chr3:180475148-181295605 (PPA4 = 0.864 and 0.862, respectively), and its DNAm levels were



correlated with the expression of *FXR1* in placenta (FDR  $P = 0.027$  and  $R^2 = 0.227$ ) (Supplementary Figure 9). *FXR1* promotes TB migration and its mRNA levels have been reported to be decreased in TB from women with recurrent spontaneous abortions<sup>51</sup>. Additionally, *FXR1* regulation of interneurons in the prefrontal cortex is critical for SCZ-like behaviors<sup>52</sup>.

Finally, we performed a Reactome gene-set analysis<sup>53</sup> of the 19 genes at the intersection of the three approaches in SCZ. We observed an enrichment for 10 gene-sets (Supplementary Data 8), including immune system pathways such as interferon alpha/beta signaling, interferon gamma signaling, and interferon signaling, supporting the idea of MIA as a link between placenta and SCZ risk.

### **Conditional SMR analysis of placental *cis*-mQTLs, and BIP, MDD and SCZ**

The developers of the SMR software designed the HEIDI test assuming a single causal variant in the *cis*-mQTL region that affects both DNAm and the trait analyzed<sup>44</sup>. 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<sup>54</sup> conditioning for the top associated *cis*-mQTL in both the GWAS and mQTL data sets. If there was a secondary signal, pointed by the presence of heterogeneity ( $P_{\text{HEIDI}} < 0.05$ ) in the *cis*-mQTL region, in either the GWAS or the mQTL dataset, we performed another round of conditional analyses conditioning only on the secondary signal in both the GWAS and mQTL datasets, 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 CpGs that passed the SMR test but failed to pass the HEIDI test. Nevertheless, due to power constraints, only those CpGs that colocalized with one of the three studied disorders according to *coloc* were considered.

We discovered a secondary hit in SCZ, in which two independently associated SNPs, i.e. rs3132386 (the primary signal) and rs16897420 (the secondary signal) had a pleiotropic effect on both cg15026241 and SCZ (Bonferroni  $P_{\text{SMR}} = 0.022$  and  $P_{\text{SMR}} = 1.67 \times 10^{-4}$ , respectively) (Supplementary Data 5 and 9). This CpG, located in the HLA region, is correlated with the expression of *ZSCAN23* in fetal placenta (FDR  $P = 0.002$  and  $R^2 = -0.282$ ) (Figure 5 and Supplementary Data 7). A strictly intergenic transcript downstream of this gene has already been identified in another SCZ study using a conditional GWAS/eQTL analysis, but it has not been characterized in detail<sup>55</sup>. Our results, namely the presence of two independent signals pointing to a single DNAm site that seems to affect gene expression in fetal placenta, together with the fact that the p-values of those SNPs do not reach genome-wide significance in the original SCZ-GWAS, strongly support causality.

### **SMR analysis of placental cell type- and GA-imQTLs in BIP, MDD and SCZ**

We used SMR to combine imQTLs with the BIP, MDD and SCZ GWAS, and obtained two STB-imQTLs (one each in BIP and SCZ), five TB-imQTLs (two in BIP, one in MDD and two in SCZ) and two GA-imQTLs (one each in BIP and SCZ) (Supplementary Data 10 and Supplementary Figure 10). Remarkably, we found an imQTL involving cg27130493 and rs72743436 (Figure 6). cg27130493 appeared to be pleiotropically associated with BIP while it changed its methylation levels as a function of STB proportion, in a rs72743436 genotype-dependent manner. This CpG is located on the gene body of *SMAD3*. The placental overexpression of this gene has been described to activate the ability of TB to form endothelial-like networks, while its defect has been associated with pre-eclampsia<sup>56</sup>. Additionally, it is a well-known target for lithium treatment in BIP<sup>57</sup>. Finally, rs72743436 is not directly associated with BIP in the original GWAS, suggesting a possible causal association between placental DNAm and BIP at this CpG site. Another interesting imQTL-CpG was cg04402684, located on a CTCF binding region, 32 kb upstream of the *HLA-E* gene. This CpG was significant in the SMR analyses performed

with TB-imQTLs in both BIP and SCZ, revealing a common DNAm site associated to both disorders (Supplementary Figure 10).

### **Comparison with SMR results for brain mQTLs**

Finally, we aimed to ascertain the tissue specificity of our findings, and thus we performed the abovementioned 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<sup>58</sup>. Briefly, mQTLs were calculated in 166 human fetal brain samples (56-166 days post-conception). The methylation data was obtained with the Infinium HumanMethylation 450K array from Illumina, and the mapping of the mQTLs was performed with the MatrixEQTL R package<sup>59</sup>, considering a *cis*-window of 0.05 Mb. Only mQTLs with a  $P < 1.5 \times 10^{-9}$  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.<sup>60</sup>, in which *cis*-mQTLs from three different datasets were meta-analyzed: in particular, 468 brain cortical region samples from the ROSMAP study<sup>61</sup>, 166 fetal brain samples by Hannon et al., and 526 frontal cortex region samples<sup>62</sup> 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 dataset was downloaded from the SMR data portal.

In BIP we obtained 13 and 35 significant SMR hits (Bonferroni  $P_{\text{SMR}} < 0.05$  and  $P_{\text{HEIDI}} > 0.05$ ) in the fetal brain and brain mQTL datasets, respectively (Supplementary Data 11 and Supplementary Figure 11), and with same criteria, we obtained 10 and 23 SMR hits in MDD, and 50 and 188 in SCZ. As shown in Figure 7, 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. In the case of the fetal brain database, the smaller sample size limits the number of detected mQTLs and could lead to underestimate the proportion of the genetic risk that acts through 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 BIP or MDD presented any overlap of DNAm sites, while there were two common hits for SCZ. cg03172226, the placental eQTM for *IRF3*, was also identified in the brain dataset in combination with rs7251, the same SNP identified in placenta. On the other hand, cg21745287, the placental eQTM for *VPS37B*, overlaps with the fetal brain results, although in combination with a different SNP than in placenta. Therefore, we cannot rule out that mostly in the case of these non-specific shared DNAm sites, the placenta could be mirroring what occurs in the brain without necessarily being the main effector tissue.

## **Discussion**

In this study, we have constructed three different placental *cis*-mQTL databases in 368 placenta samples from the INMA project: the nominal placental *cis*-mQTL database, in which all SNP-CpG combinations over a nominal p-value cutoff of  $5 \times 10^{-8}$  in 0.5 Mb windows have been compiled; the permuted placental *cis*-mQTL database, in which we have corrected for multiple correlated variants considering all the possible combinations in the same windows as above, and the conditional placental *cis*-mQTL database, that uses the permuted mQTLs to perform a stepwise regression procedure that enables to conditionally map independent *cis*-mQTLs. Importantly, we have made all results publicly available both in their raw formats and by means of a user-friendly shiny-based genome browser that enables 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.

In general, the placental *cis*-mQTLs in all three databases were depleted in regions that are usually hypomethylated, such as promoters and CpG islands, very 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, is enriched in placenta-specific active chromatin marks, and in developmental disorders of mental health and psychotic disorders, including SCZ. 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<sup>22</sup>, encouraged us to conduct a multi-omics study to try to unveil the placental contribution on the developmental origins of different neuropsychiatric disorders.

Besides, Kim-Hellmuth et al. described cell type-interacting expression QTLs (ieQTLs) to be enriched in GWAS signals and to improve GWAS-eQTL matching for the mechanistic understanding of these loci<sup>23</sup>. In this context, we wondered whether this could also be true for placental mQTLs and calculated STB- and TB-imQTLs. STB were considered for several reasons, 1) they are the most abundant cell type in term placenta<sup>63</sup>, 2) they cover the entire surface of villous trees in the placenta, and thus are in direct contact with maternal blood<sup>64</sup>, 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<sup>65,66,67,68</sup>. Given that TB are STB progenitor cells, and that there is a very high correlation between the proportions of these 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 considerable overlap and 46.3% of STB-imQTLs were also TB-imQTLs. As expected, the allelic effects in the overlapping imQTLs were negatively correlated between the two cell type-specific models, with imQTLs positively correlated with STB content being more common. This could reflect that negative interactions with a given cell type more likely reflect interactions with another cell population that is decreasing, rather than a truly negative, genotype-dependent correlation. As TB are known to differentiate into STB throughout the gestation<sup>30</sup>, 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.

Remarkably, placental DNAm is pleiotropically associated with BIP, MDD, and in particular with SCZ, while it does not seem to 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. However, very importantly, a recent article found out 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 annotatable<sup>69</sup>. In contrast, ADHD, BIP and SCZ showed the highest discoverability and hence a more annotatable 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 lack of 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 annotatability 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 similar developmental stages<sup>70</sup>.

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 early-life complications are present, the polygenic risk score (PRS)-explained risk of SCZ is more than five times higher than when they are absent<sup>71</sup>. SCZ loci that interact with early-life complications are not only highly expressed in placenta, but also differentially expressed between complicated and normal pregnancies in this organ, and enriched in response-to-stress pathways. More recently, these placenta-specific SCZ-PRS have also been shown to interact with both brain volume and cognitive function, suggesting particular neurodevelopmental trajectories in the path towards SCZ<sup>72</sup>. Particularly in the latter work, they also studied the existence of placenta-specific PRS in other neuropsychiatric disorders, and if present, whether they interacted with early neurodevelopmental outcomes. BIP presented a PRS with a very high enrichment in genes that are expressed in placenta, although this placenta-specific PRS did not interact with neither brain volume nor cognition. This makes sense since SCZ and BIP present the highest genetic correlation compared to any other pair of psychiatric traits<sup>73</sup>, but with remarkable differences in outcome, due to the fact that SCZ patients have been described to be more prone to suffer from severe cognitive impairment than BIP patients<sup>74</sup>. 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, to other neuropsychiatric disorders. This hypothesis will need to be supplemented with additional research.

Regarding the genes and pathways affected by placental DNAm pleiotropically associated to SCZ, we want to highlight that the 19 genes resulting from our multi-omics approach were enriched in immune-related terms, including interferon signaling. *HLA-H*, *IRF3* and *VPS37B* were some of the most relevant genes with immune functions. This reinforces the idea of MIA being implicated in the neurodevelopmental origins of SCZ. During pregnancy, environmental insults such as early-life complications and 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<sup>75,76,77</sup>. The MIA hypothesis proposes that exposure to a dysregulated maternal immune milieu *in utero* affects fetal neurodevelopment<sup>78,79</sup>. Moreover, in humans, maternal factors implicated in MIA are associated with epigenetic modifications in placenta<sup>80,81</sup>. 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<sup>82</sup>.

Taken individually, the hits that are most likely to be causally involved in the different neuropsychiatric disorders studied were *LRFN5* in MDD and *ZSCAN23* 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<sup>48</sup>. Additionally, it is well known that *LRFN5* is involved in the communication among brain cells and is located in a large and complex genomic niche that is highly conserved in mammals<sup>83</sup>. Specific locus structure of this region increases ASD susceptibility in males. But what makes *LRFN5* especially notable in the context of our study is the fact that we have found regional pleiotropy, that is, two independently associated CpGs in its promoter that are correlated with its expression levels in fetal placenta. These two independent signals, together with the DNAm-expression correlation in our organ of interest make the association more likely causal.

In the case of *ZSCAN23*, in a work published in 2019, conditional analyses were performed to partition the SCZ GWAS-associated brain eQTLs to better identify the downstream molecular features of genetic risk<sup>55</sup>. The top GWAS risk variant rs1233578 was associated with a strictly intergenic sequence downstream of *ZSCAN23*, and the authors claimed to have discovered a novel transcript potentially involved in SCZ through its expression in brain. This is not necessarily contradictory with our results. The function of *ZSCAN23* is unknown, although a recent paper has reported that two CpGs in the gene could show differentially variable methylation in placenta between asthmatic and control mothers<sup>84</sup>. In the present work, two SNPs have been found to independently associate with the same CpG that is related to the expression of *ZSCAN23* in placenta. Moreover, the GWAS p-values of the association between the two SNPs and SCZ are not significant. This makes placental DNAm and *ZSCAN23* expression potentially causal in SCZ, although the function of the gene and the pathways through which it may act remain unknown.

In this context, the placental DNAm sites that resulted from the SMR analyses confronting the BIP, MDD and SCZ GWAS, and the STB-, TB-, and GA-imQTLs are worthy of mention. For example, we found out that cg27130493 is pleiotropically associated with BIP while it changes its methylation levels as a function of STB proportion, in a rs72743436 genotype-dependent manner. This CpG is located in the gene body of *SMAD3*, a gene with singular functions in placenta and also in BIP itself<sup>56,57</sup>. Additionally, this 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. However, the fact that the CpG sites change not only depending on the genotype of adjacent SNPs, but also as a function of estimated placenta-specific cell content, increases the probability of 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 the different tissues studied, suggesting that the majority of our pleiotropically associated DNAm sites could be relatively specific of placenta (or at least, not related to brain DNAm). Nevertheless, *IRF3*, one of the most promising candidates due to its immune function, overlaps with the brain database. We cannot rule out neither that placenta is mirroring the brain with no effector role, nor that *IRF3* has a role in SCZ through different tissues, organs and developmental stages. Finally, 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. Especially 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 our nearly 8 million. 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<sup>72</sup>, 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 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 a part of the genetic risk to suffer from BIP, MDD and very especially SCZ, could act through placental DNAm at specific genomic loci. 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.

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, regional pleiotropic DNAm associated to the same disorder, and involvement of cell type-specific imQTLs, that additionally associate to 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.

## **Methods**

### **Placental biopsies and DNA extraction**

In the INMA<sup>24</sup> project, 2,506 mother-fetus pairs were followed until birth and a selection of 397 placentas were collected, representing the three geographical areas involved in the study. Collected placentas were stored at -80°C in a central biobank until processing. Biopsies of approximately 5 cm<sup>3</sup> were obtained from the inner region of the placenta, approximately 1.0-1.5 cm below the fetal membranes, corresponding to the villous parenchyma, and at a distance of ~5 cm from site of cord insertion. 25 mg of placental tissue was used for DNA extraction, previously rinsed twice during 5 minutes in 0.8 mL of 0.5X PBS in order 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 on a NanoDrop spectrophotometer (Thermo Scientific, Waltham, MA, USA) and additionally, 100 ng of DNA was run on 1.3% agarose gels to confirm that samples did not present visual signs of degradation. Isolated genomic DNA was stored at -20°C until further processing.

### **Genotype data**

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 the standard recommendations<sup>85,86,87</sup>. All plink files were initially processed with Will Rayner's preparation Perl script available from Mark McCarthy's Group as recommended in the documentation from the Michigan Imputation Server, using the Haplotype Reference Consortium (HRC) r1.1 2016 reference panel<sup>88</sup>. 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  $1 \times 10^{-6}$  were removed. Samples with discordant sex, those with average heterozygosity values above or below 4 standard deviations or with more than 3% missing genotype 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 dataset was imputed with the Michigan Imputation Server<sup>88</sup> using the HRC reference panel, Version r1.1 2016<sup>89</sup>. Before imputation, data was converted into VCF format. Phasing of haplotypes was done with Eagle v2.4<sup>90</sup> and genotype imputation with Minimac4<sup>91</sup>, both implemented in the code by the Michigan Imputation Server. Finally, we removed variants with an imputation dosage  $r^2$  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 paired methylation data were considered in this analysis. The final dataset consisted in 368 samples and 4,171,035 SNPs.

### **Methylation data**

DNA methylation was assessed with the Infinium MethylationEPIC BeadChip from Illumina, following manufacturer's protocol in the Erasmus Medical Centre core facility. Briefly, 750 ng of DNA from 397 placental samples were bisulphite-converted using the EZ 96-DNA methylation kit from Zymo Research, following the manufacturer's standard protocol, and DNA methylation was measured using the Infinium protocol. Three technical duplicates were included. Samples were randomized

taking into account region-of-origin and sex. As the number of samples in each condition was different, a perfect randomization was not possible. However, all the plates had samples from all three involved geographical areas, and an equilibrated number of male and female samples.

The quality control of the methylation data, including 865,859 DNAm probes, was performed using the PACEAnalysis R package (v.0.1.7)<sup>92</sup>. Before starting the quality control with PACEAnalysis, one sample was discarded for having 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 exhibiting contamination with DNA from another subject or the mother. Only those samples with paired genotype data were considered in this study. Probes with a call rate smaller than 95%, within sexual chromosomes, SNPs (European MAF < 5%) and cross-hybridizing potential were excluded from the analysis.

The methylation beta values were normalized in different steps. Dye-bias and Noob background correction, implemented in minfi R package, were applied<sup>93,94</sup>, followed by normalization of the data with the functional normalization method<sup>95</sup>. Then, to correct for the bias of type-2 probes values the beta-mixture quantile (BMIQ) normalization was applied<sup>96</sup>. After that, we explored the clustering of the data through Principal Component Analysis (PCA) and tested the association of the 12 first PCs with the main and the technical variables. Array batch effect was controlled with the ComBat method<sup>97</sup>. Finally, to correct for the possible outliers, we winzorized the extreme values to the 1% percentile (0.5% in each side), where percentiles were estimated with the empirical beta-distribution. The final dataset consisted in 368 samples and 747,486 DNAm probes (CpGs).

Cell type proportions of six populations (STB, TB, nucleated red blood cells, Hofbauer cells, endothelial cells, and stromal cells) were estimated from DNAm using the placenta reference panel from the 3<sup>rd</sup> trimester implemented in the Planet R package<sup>31</sup>.

### **Placental *cis*-mQTL analysis**

A total amount of 4,171,035 SNPs, 747,486 CpGs and 368 samples with paired genotype and DNAm data were considered for the *cis*-mQTL analysis in TensorQTL<sup>98</sup>. TensorQTL nominal modality performs linear regressions between the genotype and the normalized DNAm beta-values, as implemented in FastQTL<sup>99</sup>. The covariates included in the regression model were the sex of the fetuses, the first five PCs derived from the genotype data (genotype PCs) and the five cell type proportions derived from the Planet methylation panel. Genotype PCs were included in the model as covariates to remove the hidden batch effects and other potential confounders in the genotype data. The TensorQTL *cis*-region was specified as  $\pm 0.5$  Mb from each tested CpG position, consistent with the results of previous studies where the distance between the majority *cis*-mQTL SNPs and DNAm sites is  $\leq 0.5$  Mb<sup>59,100,101,102</sup>. In the final nominal *cis*-mQTL database we included only those probes with at least one *cis*-mQTL at  $P_{\text{nominal}} < 5 \times 10^{-8}$ . This same regression model was used to build 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 it is described in Ongen et al. 2016<sup>99</sup>. Conversely, 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<sup>103</sup>.

### **Characterization of placental *cis*-mQTLs**

The separate characterization of the nominal, permuted and conditional mQTL databases was performed in different steps. First, the distribution of both, the distance between mQTL-CpGs and



paired mQTL-SNPs, and the mQTLs along the chromosomes was plotted, followed by a volcano plot to check the uniform distribution of the negative and positive effect sizes.

Second, using the annotation from IlluminaHumanMethylationEPICanno.ilm10b4.hg19 R package<sup>104</sup>, we performed several Chi-Square tests to assess the enrichment and depletion of the UCSC RefGene and Relation to CpG Island annotations. We also depicted density plots of the methylation values according to the Relation to CpG Island annotation. Moreover, for the top 10,000 mQTL-CpGs, we assessed enrichment and depletion of 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<sup>105</sup> using eFORGE v2.0<sup>26,27,28</sup>. The enrichment and depletion with each of the three putative functional elements were tested separately and compared to the respective data from the consolidated ROADMAP epigenomics reference panel.

Lastly, over-representation analyses of Gene Ontology (GO) and Kyoto Encyclopaedia of Genes and Genomes (KEGG) gene sets were conducted with MissMethyl R package<sup>106,107,108,109</sup>. MissMethyl performs a hypergeometric test taking into account the bias derived from the differing number of probes per gene and the multiple genes annotated per CpG<sup>106</sup>. Gene set enrichment analyses with the Disease Ontology (DO) database were conducted with the DOSE R package<sup>110</sup>. With DOSE, we performed the gene set enrichment analysis considering the genes annotated in the IlluminaHumanMethylationEPICanno.ilm10b4.hg19 R package<sup>104</sup> as background.

### Placental *cis*-imQTL analysis

The cell type-imQTLs were computed with the interaction modality from TensorQTL, consisting in nominal associations for a linear model that includes a genotype per interaction term<sup>98</sup>. The same initial SNPs, CpGs and samples from the nominal database were considered for the interaction analysis. The covariates used were the sex of the fetuses and the first five genotype PCs. Following the recommendations by Kim-Hellmuth S. et al.2021<sup>23</sup>, estimations of STBs and TBs per sample were defined separately as the interaction terms. In the two final interaction *cis*-mQTL databases, we included only probes with at least one *cis*-mQTL at  $P_{\text{nominal}} < 5 \times 10^{-8}$ . The imQTLs obtained from each model were categorized as positively and negatively correlated with cell type estimates, or as uncertain if this was ambiguous. To categorize the imQTLs into these groups, genotype main effects at low (<25<sup>th</sup> percentile) vs high (>75<sup>th</sup> percentile) cell type proportion were compared. imQTLs with positive cell type correlation showed an increase of the genotype main effect from low to high cell proportions ( $\beta_{\text{mQTL low}} < \beta_{\text{mQTL high}}$ ). imQTLs with negative cell type correlation showed a decrease ( $\beta_{\text{mQTL low}} > \beta_{\text{mQTL high}}$ ) and the uncertain group contained imQTLs where the sign flipped between low and high cell type proportions ( $\beta_{\text{mQTL low}} \neq \beta_{\text{mQTL high}}$ ).

Additionally, GA was also considered as interaction term. In this case, the covariates considered were the sex of the fetuses, the first five genotype PCs and the cell type proportions from Planet.

### Genome Wide Association Studies

The GWAS summary-statistics used in this analysis were the public largest to-date association studies of ADHD<sup>32</sup>, AGR<sup>33</sup>, ASD<sup>34</sup>, BIP<sup>35</sup>, INT<sup>36</sup>, MDD<sup>37</sup>, OCD<sup>38</sup>, PD<sup>39</sup>, SA<sup>40</sup>, and SCZ<sup>41</sup> from EAGLE, Indonesia Schizophrenia Consortium, International Obsessive Compulsive Disorder Foundation Genetics Collaborative (IOCDF-GC), OCD Collaborative Genetics Association Studies (OC GAS), PsychENCODE, Psychiatric Genomics Consortium (PGC), Psychosis Endophenotypes International Consortium, UK Biobank (UKB), SynGO Consortium, and 23andMe, among others.

All the GWAS summary-statistics were harmonized according to the dbSNP build 155 and the INMA genotype data as a reference. The harmonization steps included: to change rsID to chromosome:position nomenclatures, to correct the effect allele, the effect size and the allele frequency according to the reference genotypes if applicable, and to create a .ma format file with the summary-statistics as indicated in the SMR pipeline<sup>44</sup>.

### Multi-SNP-based Mendelian Randomization analysis

Mendelian Randomization analysis was carried out considering *cis*-mQTL SNPs as the instrumental variables (IVs), CpG methylation as the exposure (X), and the neuropsychiatric traits as the outcome (Y). Multi-SNP-based MR (SMR-multi)<sup>111</sup> analysis was performed by the SMR software<sup>43</sup>. SMR integrates GWAS and QTL summary-statistics to test for pleiotropic associations between quantitative traits, such as methylation or expression, and a complex trait, for instance 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  $P_{\text{nominal}} < 5 \times 10^{-8}$  in the *cis* region (0.5 Mb of the CpG). Second, it removed SNPs in very high 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<sup>111</sup>. 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<sup>45</sup>, 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 ( $P_{\text{nominal}} < 5 \times 10^{-8}$ ) 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  $P_{\text{SMR}}$  corrected Bonferroni  $< 0.05$  and  $P_{\text{HEIDI}} > 0.05$  (not showing heterogeneity).

### Colocalization analyses

Recalling the definition of *cis*-mQTLs in our analysis, all DNAm sites with at least one significant mQTL ( $P_{\text{nominal}} < 5 \times 10^{-8}$ ) and located within 0.5 Mb of the regions defined in the original neuropsychiatric GWAS were tested. Colocalization analysis was performed as described in Giambartolomei C. et al.2014 with the R *coloc* package<sup>45</sup>. In total 17,341, 5,466, and 4,643 mQTL-CpGs were tested for BIP ( $n = 63$  autosomal regions), MDD ( $n = 98$  autosomal regions) and SCZ ( $n = 279$  autosomal regions), respectively. In both the GWAS data and our mQTLs we imputed the regression coefficients, their variances and the SNP minor allele frequencies, and the prior probabilities were left as their default values. This methodology quantifies the support across the results of each GWAS for five hypotheses by calculating the posterior probabilities, denoted as  $PP_i$  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, DNA methylation;

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

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

### Conditional analyses

We performed a GCTA conditional analysis<sup>112,113</sup> conditioning on the top *cis*-mQTL of those probes that passed SMR ( $P_{\text{SMR}}$  Bonferroni  $< 0.05$ ) but failed to pass the HEIDI test ( $P_{\text{HEIDI}} < 0.05$ ), 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 (BP, MDD, SCZ). For any of these regions where there was evidence of a secondary signal ( $P_{\text{conditional}} < 5 \times 10^{-8}$ ), 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-corrected  $P_{\text{SMR}} < 0.05$  and  $P_{\text{HEIDI}} > 0.05$  (not showing heterogeneity).

### **RICHs eQTLs**

RICHs recruited mother and infant pairs from March 2009 to May 2013 following delivery at the Women and Infants Hospital of Rhode Island. RICHs selected infants both small for GA, large for GA and controls born appropriate for GA matched on sex, GA ( $\pm 3$  days), and maternal age ( $\pm 2$  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<sup>114</sup>. About 20 million single-end RNA-seq reads were generated on each sample<sup>115</sup>. Placental DNAm data ( $n = 220$ ) were obtained using the Illumina Infinium HumanMethylation450 BeadChip, preprocessed, and normalized with the minfi R package<sup>116</sup>. eQTLs were calculated by implementing linear models in MatrixEQTL, considering *cis*-windows of 0.5 Mb up and downstream of each CpG in a total of 195 placenta samples. Linear regressions were adjusted by sex, 5 PCs of expression and the Planet estimated cell types. Results were corrected with FDR.

The gene set enrichment analysis of the eQTL-genes associated to SCZ according to our SMR and colocalization approaches was performed using the Reactome gene list analysis<sup>53</sup>, with the default settings recommended by the developers in the Reactome database release 83 and the pathway browser version 3.7.

### **Code availability**

The code for the genotype and methylation QC, as well as 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 analysis is available in this GitHub repository link: <https://github.com/ariadnacilleros/Cilleros-PortetA.etal>.

### **Data availability**

The three complete placental *cis*-mQTL databases are publicly available online in the following address: <https://smari.shinyapps.io/shi>. To be uploaded to a repository.

## Bibliography

1. Strauss, R.S. Effects of the intrauterine environment on childhood growth. *Br. Med. Bull.* 53, 81-95 (1997).
2. Barker, D.J. The fetal and infant origins of adult disease. *BMJ.* 301, 1111 (1990).
3. O'Donnell, K., O'Connor, T. G., & Glover, V. Prenatal stress and neurodevelopment of the child: focus on the HPA axis and role of the placenta. *Dev. Neurosci.* 31, 285–292 (2009).
4. Lapehn, S., Paquette, A.G. The Placental Epigenome as a Molecular Link Between Prenatal Exposures and Fetal Health Outcomes Through the DOHaD Hypothesis. *Curr. Environ. Health. Rep.* 9, 490-501 (2022).
5. Lester, B.M., Marsit, C.J. Epigenetic mechanisms in the placenta related to infant neurodevelopment. *Epigenomics* 10, 321-333 (2018).
6. Yen, S.S. The placenta as the third brain. *J. Reprod. Med.* 39, 277-280 (1994).
7. Rapoport, J.L., Addington, A.M., Frangou, S., Psych, M.R. The neurodevelopmental model of schizophrenia: update 2005. *Mol. Psychiatry* 10, 434-449 (2005).
8. Murray, R.M., Lewis, S.W. Is schizophrenia a neurodevelopmental disorder? *Br. Med. J.* 295, 681-682 (1987).
9. Owen, M.J., O'Donovan, M.C. Schizophrenia and the neurodevelopmental continuum: evidence from genomics. *World Psychiatry* 16, 227-235 (2017).
10. Weinberger, D.R. The neurodevelopmental origins of schizophrenia in the penumbra of genomic medicine. *World Psychiatry* 16, 225-226 (2017).
11. Canetta, S.E., Brown, A.S. PRENATAL INFECTION, MATERNAL IMMUNE ACTIVATION, AND RISK FOR SCHIZOPHRENIA. *Transl. Neurosci.* 3, 320-327 (2012).
12. Carnac, T. Schizophrenia Hypothesis: Autonomic Nervous System Dysregulation of Fetal and Adult Immune Tolerance. *Front. Syst. Neurosci.* 16, 844383 (2022).
13. Novakovic, B., Saffery, R. The ever growing complexity of placental epigenetics - role in adverse pregnancy outcomes and fetal programming. *Placenta* 33, 959-970 (2012).
14. Konwar, C., Del Gobbo, G., Yuan, V., Robinson, W.P. Considerations when processing and interpreting genomics data of the placenta. *Placenta* 84, 57-62 (2019).
15. Loyfer, N., et al. A DNA methylation atlas of normal human cell types. *Nature* 613, 355-364 (2023).
16. Everson T.M., et al. Placental DNA methylation signatures of maternal smoking during pregnancy and potential impacts on fetal growth. *Nat. Commun.* 12, 5095(2021).
17. Fernandez-Jimenez N., et al. A meta-analysis of pre-pregnancy maternal body mass index and placental DNA methylation identifies 27 CpG sites with implications for mother-child health. *Commun. Biol.* 5, 1313(2022).
18. Do, C., et al. Mechanisms and Disease Associations of Haplotype-Dependent Allele-Specific DNA Methylation. *Am. J. Hum. Genet.* 98, 934-955(2016).
19. Delahaye, F., et al. Genetic variants influence on the placenta regulatory landscape. *PLoS Genet.* 15, e1008118(2019).
20. Tekola-Ayele, F., et al. Placental multi-omics integration identifies candidate functional genes for birthweight. *Nat. Commun.* 13, 2384(2022).
21. Casazza, W., et al. 2022. Sex-dependent placental mQTL provide insight into the prenatal origins of childhood-onset traits and conditions. *medRxiv* doi: 10.1101/2022.10.04.22280695 (2022).

22. Oliva, M., et al. DNA methylation QTL mapping across diverse human tissues provides molecular links between genetic variation and complex traits. *Nat. Genet.* 55, 112-122(2023).
23. Kim-Hellmuth, S., et al. Cell type-specific genetic regulation of gene expression across human tissues. *Science* 369, eaaz8528 (2020).
24. Guxens, M., et al. Cohort Profile: The INMA—INfancia y Medio Ambiente—(Environment and Childhood). *Proj. Int. J. Epidemiol.* 41, 930–940 (2012).
25. Yuan, V., Hui, D., Yin, Y., Peñaherrera, M. S., Beristain, A. G., Robinson, W. P. Cell-specific characterization of the placental methylome. *BMC Genomics* 22, 6 (2021).
26. Min, J.L., et al. Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. *Nat. Genet.* 53, 1311-1321 (2021).
27. Breeze, C.E., et al. eFORGE v2.0: updated analysis of cell type-specific signal in epigenomic data. *Bioinformatics* 35, 4767-4769 (2019).
28. Breeze, C.E., et al. eFORGE: A Tool for Identifying Cell Type-Specific Signal in Epigenomic Data. *Cell. Rep.* 17, 2137-2150 (2016).
29. Breeze, C.E. Cell Type-Specific Signal Analysis in Epigenome-Wide Association Studies. *Methods in Molecular Biology* (Clifton, N.J.). 2432, 57-71 (2022).
30. Schriml, L.M., et al. Human Disease Ontology 2018 update: classification, content and workflow expansion. *Nucleic Acids Res.* 47, D955-D962 (2019).
31. Ellery, P.M., Cindrova-Davies, T., Jauniaux, E., Ferguson-Smith, A.C., Burton, G.J. Evidence for transcriptional activity in the syncytiotrophoblast of the human placenta. *Placenta* 30, 329-334.
32. Demontis, D., et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat. Genet.* 51, 63-75 (2019).
33. Pappa, I., et al. A genome-wide approach to children's aggressive behavior: The EAGLE consortium. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* 171, 562-572 (2016).
34. 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. *Mol. Autism.* 8, 21 (2017).
35. Mullins, N., et al. Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nat. Genet.* 53, 817-829 (2021).
36. Jami, E.S., et al. Genome-wide Association Meta-analysis of Childhood and Adolescent Internalizing Symptoms. *J. Am. Acad. Child Adolesc. Psychiatry.* 61, 934-945 (2022).
37. Howard, D.M., et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat. Neurosci.* 22, 343-352 (2019).
38. 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. *Mol. Psychiatry.* 23, 1181-1188 (2018).
39. Forstner, A.J., et al. Genome-wide association study of panic disorder reveals genetic overlap with neuroticism and depression. *Mol. Psychiatry.* 26, 4179-4190 (2021).
40. Jiang, L., Zheng, Z., Fang, H., Yang, J. A generalized linear mixed model association tool for biobank-scale data. *Nat. Genet.* 53, 1616-1621 (2021).
41. Trubetskov, V., et al. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* 604, 502-508 (2022).

42. *Psychiatric Genomics Consortium. PGC*. Available at: <https://pgc.unc.edu/> (2007). (Accessed: 23<sup>rd</sup> January 2023).
43. *EARly Genetics and Lifecourse Epidemiology. EAGLE*. Available at: <https://eagle-consortium.org/>. (Accessed: 23<sup>rd</sup> January 2023).
44. Zhu, Z., et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* 48, 481-487 (2016).
45. Giambartolomei, C., et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* 10, e1004383 (2014).
46. Appleton, A.A., et al. Prenatal Programming of Infant Neurobehaviour in a Healthy Population. *Paediatr. Perinat. Epidemiol.* 30, 367–375 (2016).
47. Qi, X., et al. An integrative analysis of genome-wide association study and regulatory SNP annotation datasets identified candidate genes for bipolar disorder. *Int. J. Bipolar Disord.* 8, 6 (2020).
48. 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. *Transl. Psychiatry.* 9, 310 (2019).
49. Mischler, A., et al. Two distinct trophoblast lineage stem cells from human pluripotent stem cells. *J. Biol. Chem.* 296, 100386 (2021).
50. Ben-Yehuda, H., et al. Maternal Type-I interferon signaling adversely affects the microglia and the behavior of the offspring accompanied by increased sensitivity to stress. *Mol. Psychiatry.* 25, 1050-1067 (2020).
51. Hong, W., Chen, J.H., Ma, H.J., Li-Li, Li, X.C. Fragile X-Related Protein 1 (FXR1) Promotes Trophoblast Migration at Early Pregnancy via Downregulation of GDF-15 Expression. *Reprod. Sci.* 29, 110-121 (2022).
52. Shen, M., et al. FXR1 regulation of parvalbumin interneurons in the prefrontal cortex is critical for schizophrenia-like behaviors. *Mol. Psychiatry.* 26, 6845-6867 (2021).
53. Gillespie, M., et al. The reactome pathway knowledgebase 2022. *Nucleic Acids Res.* 50, D687-D692 (2022).
54. Yang, J., et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* 44, 369–375 (2012).
55. Jaffe, A.E., et al. Developmental and genetic regulation of the human cortex transcriptome illuminate schizophrenia pathogenesis. *Nat. Neurosci.* 21, 1117-1125 (2018).
56. Brkić, J., et al. Differential Role of Smad2 and Smad3 in the Acquisition of an Endovascular Trophoblast-Like Phenotype and Preeclampsia. *Front. Endocrinol. (Lausanne)* 11, 436 (2020).
57. Liang, M.H., Wendland, J.R., Chuang, D.M. Lithium inhibits Smad3/4 transactivation via increased CREB activity induced by enhanced PKA and AKT signaling. *Mol. Cell. Neurosci.* 37, 440-453 (2008).
58. Hannon, E., et al. Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci. *Nat. Neurosci.* 19, 48-54 (2016).
59. Shabalín, A.A. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* 28, 1353-1358 (2012).
60. Qi, T., et al. Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. *Nat. Commun.* 9, 2282 (2018).
61. Ng, B., et al. An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. *Nat. Neurosci.* 20, 1418-1426 (2017).

62. Jaffe, A.E., et al. Mapping DNA methylation across development, genotype and schizophrenia in the human frontal cortex. *Nat. Neurosci.* 19, 40-47 (2016).
63. 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).
64. Ellery, P.M., Cindrova-Davies, T., Jauniaux, E., Ferguson-Smith, A.C., Burton, G.J. Evidence for transcriptional activity in the syncytiotrophoblast of the human placenta. *Placenta* 30, 329–334 (2009).
65. White, R., Jobling, S., Hoare, S. A., Sumpter, J. P., Parker, M.G. *Endocrinology* 135, 175 (1994).
66. McLaren, A. *J. Endocrinol.* 42,453 (1968); McLaren, A., in *Ovo-Implantation: Human Gonadotropins and Prolactin*, Hubinout, P. O., et al., Eds. (Karger, NewYork, 1970), pp.18-37.
67. Finnand, C. A., Bredl, J. C. S. *J. Reprod. Fert.* 34, 247(1973)
68. Bragaand, V. M., Gendler, S. J., *J. Cell Sci.* 105,397(1993).
69. Romero, C., Werme, J., Jansen, P. R., et al. Exploring the genetic overlap between twelve psychiatric disorders. *Nat Genet.* 54, 1795-1802 (2022).
70. Rao, S., Yin, L., Xiang, Y., So, H. C. Correction: Analysis of genetic differences between psychiatric disorders: exploring pathways and cell types/tissues involved and ability to differentiate the disorders by polygenic scores. *Transl Psychiatry.* 12, 237 (2022).
71. Ursini, G., et al. Convergence of placenta biology and genetic risk for schizophrenia. *Nat. Med.*, 24, 792-801 (2018).
72. Ursini, G., et al. Placental genomic risk scores and early neurodevelopmental outcomes. *Proc. Natl. Acad. Sci. U S A* 118, e2019789118 (2021).
73. Grotzinger, A.D., et al. Genetic architecture of 11 major psychiatric disorders at biobehavioral, functional genomic and molecular genetic levels of analysis. *Nat Genet.* 54, 548-559 (2022).
74. Bora, E. Differences in cognitive impairment between schizophrenia and bipolar disorder: Considering the role of heterogeneity. *Psychiatry Clin Neurosci.* 70, 424-433 (2016).
75. Han, V.X., Patel, S., Jones, H.F., Dale, R. C. Maternal immune activation and neuroinflammation in human neurodevelopmental disorders. *Nat Rev Neurol.* 17, 564-579 (2021).
76. Weber-Stadlbauer, U. Epigenetic and transgenerational mechanisms in infection-mediated neurodevelopmental disorders. *Transl. Psychiatry* 7, e1113 (2017).
77. Richetto, J., et al. Genome- wide DNA methylation changes in a mouse model of infection- mediated neurodevelopmental disorders. *Biol. Psychiatry* 81, 265–276 (2017).
78. Estes, M. L., McAllister, A. K. Maternal immune activation: implications for neuropsychiatric disorders. *Science* 353, 772–777 (2016).
79. Knuesel, I., et al. Maternal immune activation and abnormal brain development across CNS disorders. *Nat. Rev. Neurol.* 10, 643–660 (2014).
80. Banik, A., et al. Maternal factors that induce epigenetic changes contribute to neurological disorders in offspring. *Genes* 8, 150 (2017).
81. Hodyl, N. A., Roberts, C. T., Bianco-Miotto, T. Cord blood DNA methylation biomarkers for predicting neurodevelopmental outcomes. *Genes* 7, 117 (2016)
82. Howerton, C. L., Bale, T. L. Prenatal programming: at the intersection of maternal stress and immune activation. *Horm. Behav.* 62, 237–242 (2012)
83. Lybaek, H., et al. LRFN5 locus structure is associated with autism and influenced by the sex of the individual and locus conversions. *Autism Res.* 15, 421-433 (2022).

84. Lee, A. G., et al. Associations between antenatal maternal asthma status and placental DNA methylation. *Placenta*. 126, 184-195 (2022).
85. Anderson, C. A., Pettersson, F.H., Clarke, G.M., Cardon, L. R., Morris, A. P., Zondervan, K. T. Data quality control in genetic case-control association studies. *Nat Protoc.* 5, 1564-1573 (2010).
86. Purcell, S., et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 81, 559-575 (2007).
87. Chang, C.C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., Lee, J. J. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience.* 4, 7 (2015).
88. Das, S., et al. Next-generation genotype imputation service and methods. *Nat Genet.* 48, 1284-1287 (2016).
89. McCarthy, S., et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet.* 48, 1279-1283 (2016).
90. Loh, P. R., et al. Reference-based phasing using the Haplotype Reference Consortium panel. *Nat Genet.* 48, 1443-1448 (2016).
91. Fuchsberger, C., Abecasis, G. R., Hinds, D. A. minimac2: faster genotype imputation. *Bioinformatics* 31, 782-784 (2015).
92. Binder, A.M. QC steps: Epi-Centered Research. <https://www.epicenteredresearch.com/pace/qcsteps/> (2021).
93. Fortin, J. P., Triche, T. J. Jr., Hansen, K. D. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. *Bioinformatics* 33, 558-560 (2017).
94. Triche, T. J. Jr., Weisenberger, D. J., Van Den Berg, D., Laird, P. W., Siegmund, K. D. Low-level processing of Illumina Infinium DNA Methylation BeadArrays. *Nucleic Acids Res.* 41, e90 (2013).
95. Fortin, J. P., et al. Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biol.* 15, 503 (2014).
96. 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).
97. Johnson, W. E., Li, C., Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 8, 118-127 (2007).
98. Taylor-Weiner, A., et al. Scaling computational genomics to millions of individuals with GPUs. *Genome Biol.* 20, 228 (2019).
99. Ongen, H., Buil, A., Brown, A. A., Dermitzakis, E. T., Delaneau, O. Fast and efficient QTL mapper for thousands of molecular phenotypes. *Bioinformatics* 32, 1479-1485 (2016).
100. Gibbs, J.R., et al. Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet.* 6, e1000952 (2010).
101. 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).
102. Olsson, A.H., et al. Genome-wide associations between genetic and epigenetic variation influence mRNA expression and insulin secretion in human pancreatic islets. *PLoS Genet.* 10, e1004735 (2014).



103. 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).
104. Hansen, K. D. Illumina Human Methylation EPICanno.ilm10b4.hg19: Annotation for Illumina's EPIC methylation arrays. R package version 0.6.0, [https://bitbucket.com/kasperdanielhansen/Illumina\\_EPIC](https://bitbucket.com/kasperdanielhansen/Illumina_EPIC) (2017).
105. Roadmap Epigenomics Consortium, et al. Integrative analysis of 111 reference human epigenomes. *Nature* 518, 317-330 (2015).
106. Maksimovic, J., Gordon, L., Oshlack, A. SWAN: Subset-quantile within array normalization for illumina Infinium Human Methylation450 BeadChips. *Genome Biol.* 13, R44 (2012).
107. Phipson, B., Oshlack, A. DiffVar: a new method for detecting differential variability with application to methylation in cancer and aging. *Genome Biol.* 15, 465 (2014).
108. Maksimovic, J., Gagnon-Bartsch, J. A., Speed, T. P., Oshlack, A. Removing unwanted variation in a differential methylation analysis of Illumina Human Methylation450 array data. *Nucleic Acids Res.* 43, e106 (2015).
109. Phipson, B., Maksimovic, J., Oshlack, A. missMethyl: an R package for analyzing data from Illumina's Human Methylation450 platform. *Bioinformatics* 32, 286-288 (2016).
110. Yu, G., Wang, L. G., Yan, G. R., He, Q. Y. DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. *Bioinformatics* 31, 608-609 (2015).
111. Wu, Y., et al. Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. *Nat Commun.* 9, 918 (2018).
112. Yang, J., Lee, S. H., Goddard, M. E., Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 88, 76-82 (2011).
113. Yang, J. et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* 44, 369-375 (2012).
114. Love, M. I., Hogenesch, J. B., Irizarry, R. A. Modeling of RNA-seq fragment sequence bias reduces systematic errors in transcript abundance estimation. *Nat. Biotechnol.* 34, 1287-1291 (2016).
115. Peng, S. et al. Expression quantitative trait loci (eQTLs) in human placentas suggest developmental origins of complex diseases. *Hum. Mol. Genet.* 26, 3432-3441 (2017).
116. Fortin, J. P., Triche, T. J. Jr., Hansen, K. D. Preprocessing, normalization and integration of the Illumina Human Methylation EPIC array with minfi. *Bioinformatics* 33, 558-560 (2017).

## Acknowledgements

We would like to acknowledge all the INMA and RICHS participants and researchers, for their kind collaboration and support.

## Funding

INMA-Gipuzkoa is funded by grants from Instituto de Salud Carlos III (PI06/0867 and PI09/00090, incl. FEDER funds), Department of Health of the Basque Government (2005111093), Provincial Government of Gipuzkoa (DFG06/002), and annual agreements with the municipalities of the study area (Zumarraga, Urretxu, Legazpi, Azkoitia, Azpeitia and Beasain). INMA-Sabadell was funded by grants from Instituto de Salud Carlos III (Red INMA G03/176, PS09/00432, PI17/01225, PI17/01935 and CP18/00018), Fundació La Marató de TV3 (090430), and Generalitat de Catalunya-CIRIT (1999SGR 00241), the European Community's Seventh Framework Programme (FP7/2007-206) under grant agreement no 308333 (HELIX project), and from the European Joint Programming Initiative "A Healthy Diet for a Healthy Life" (JPI HDHL and Instituto de Salud Carlos III) under the grant agreement no AC18/00006 (NutriPROGRAM project). ISGlobal acknowledges support from the Spanish Ministry of Science and Innovation and the State Research Agency through the "Centro de Excelencia Severo Ochoa 2019-2023" Program (CEX2018-000806-S), and support from the Generalitat de Catalunya through the CERCA Program. INMA-Valencia is funded by Grants from UE (FP7-ENV-2011 cod 282957, HEALTH.2010.2.4.5-1, and H2020 No 874583, the ATHLETE project), the Ministry of Universities (CAS21/00008, Margarita Salas Grant MS21-133 and NextGeneration EU), Instituto de Salud Carlos III (FIS-FEDER: 13/1944, 16/1288, 17/00663 and 19/1338; FIS-FSE: 17/00260; Miguel Servet-FSE: MSII20/0006), CIBERESP, Generalitat Valenciana (BEST/2020/059, AICO/2020/285 and CIAICO/2021/132). The RICHS cohort is supported by the US National Institute of Environmental Health Sciences (U24 ES02507). M.C.T. is funded by a Beatriu de Pinós Postdoctoral Contract awarded by Generalitat de Catalunya-AGAUR and European Commission-Horizon 2020 (2019 BP 00107). I.G.S. is funded by the Basque Department of Health (SAN2020111043) and the Spanish Ministry of Equity (12-4-ID22) and the UPV/EHU Collaborative Projects (COLAB22/01). A.H.L. is a predoctoral fellow supported by grant PRE-C-2020-0091 from the MCIN/AEI/10.13039/501100011033 and by ESF Investing in your future. B.P.G.G. is supported by the Mexican National Council for Science and Technology grant 2021-000007-01EXTF-00209. M.F.F. is funded by the EU Commission (QLK4-1999-01422, QLK4-2002-00603 and CONTAMED FP7-ENV-212502) and the Consejería de Salud de la Junta de Andalucía (Grant number 0675/10). J.R.B. is funded by Research Grant PID2019-106382RB-I00 funded by MCIN/AEI/10.13039/501100011033. N.F.J. is funded by research grants 2019/111085 from the Basque Department of Health, and PI21/01491 from the Instituto de Salud Carlos III (ISCIII), co-funded by the European Union.

## Author contributions

N.F.J., J.R.B. and A.C.P conceived and designed the study. Data analyses were carried out by A.C.P., C.L., S.M., C.B. and N.F.J. I.G.S., D.G.M., A.H.L., B.P.G.G. and M.B. helped with the data analysis and provided valuable analytical advice. M.C.T., M.L., A.I., A.B., G.E., R.S.B., L.S.M., J.C., S.L., M.F.F., M.V., J.I., M.G., C.M., M.B. and J.R.B. actively participated in sample recruitment and/or data acquisition. A.C.P., N.F.J. and J.R.B. interpreted the results. N.F.J. and A.C.P. wrote the first draft of the manuscript. N.F.J. directed the project. All authors (A.C.P., C.L., S.M., M.C.T., M.L., A.I., A.B., I.G.S., D.G.M., G.E., A.H.L., R.S.B., C.E.B., B.P.G.G., L.S.M., J.C., S. L., M.F.F., M.V., J.I., M.G., C.M., M.B., J.R.B. and N.F.J.) made substantial contributions to the acquisition, analysis, or interpretation of data, and read and critically revised the manuscript.

## Tables

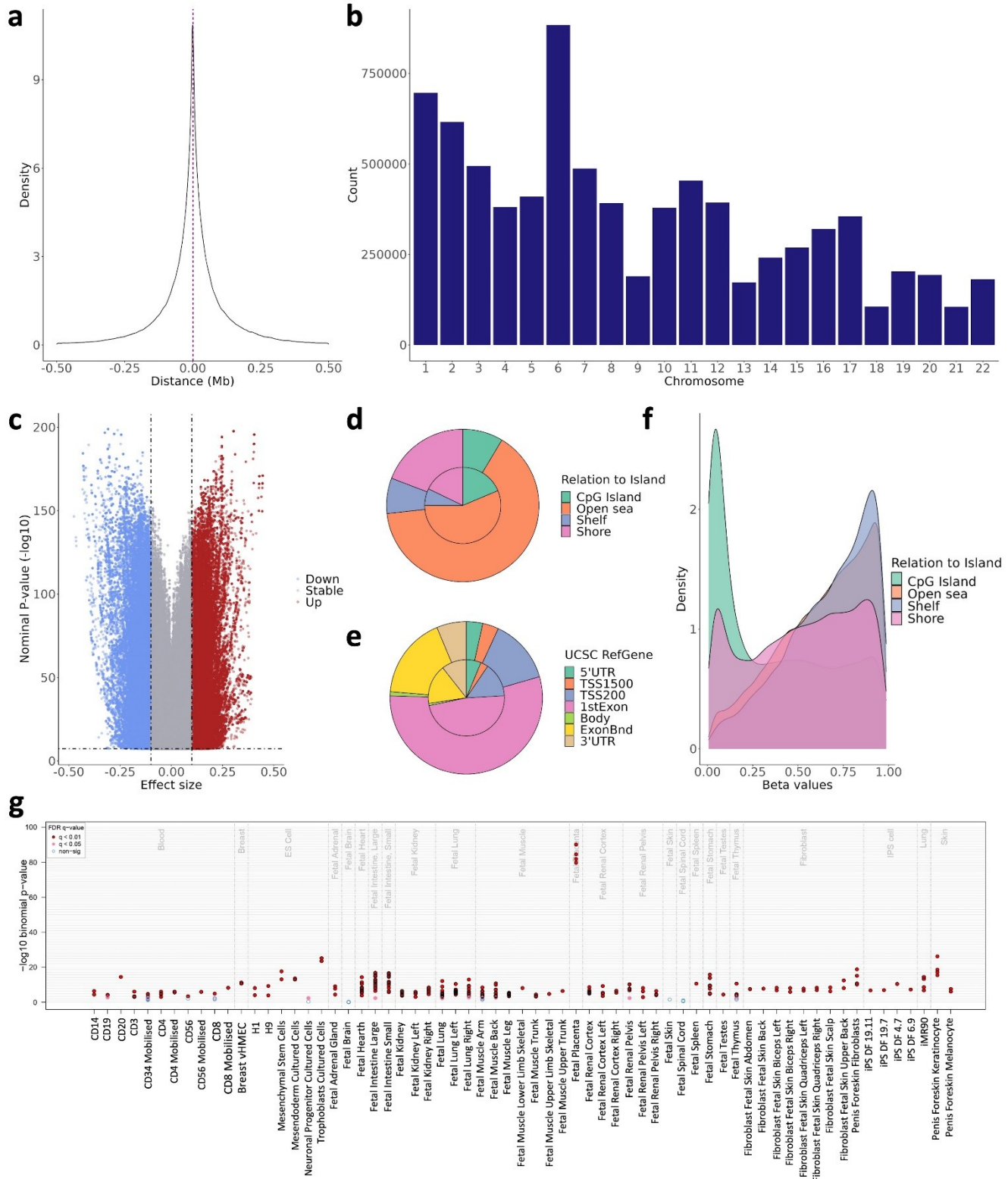
**Table 1.** Distribution of maternal smoking during pregnancy, demographic variables, birth outcomes and covariates of the INMA cohort. SD = Standard Deviation.

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

**Table 2.** Overview of the neuropsychiatric traits and disorders included in this study.

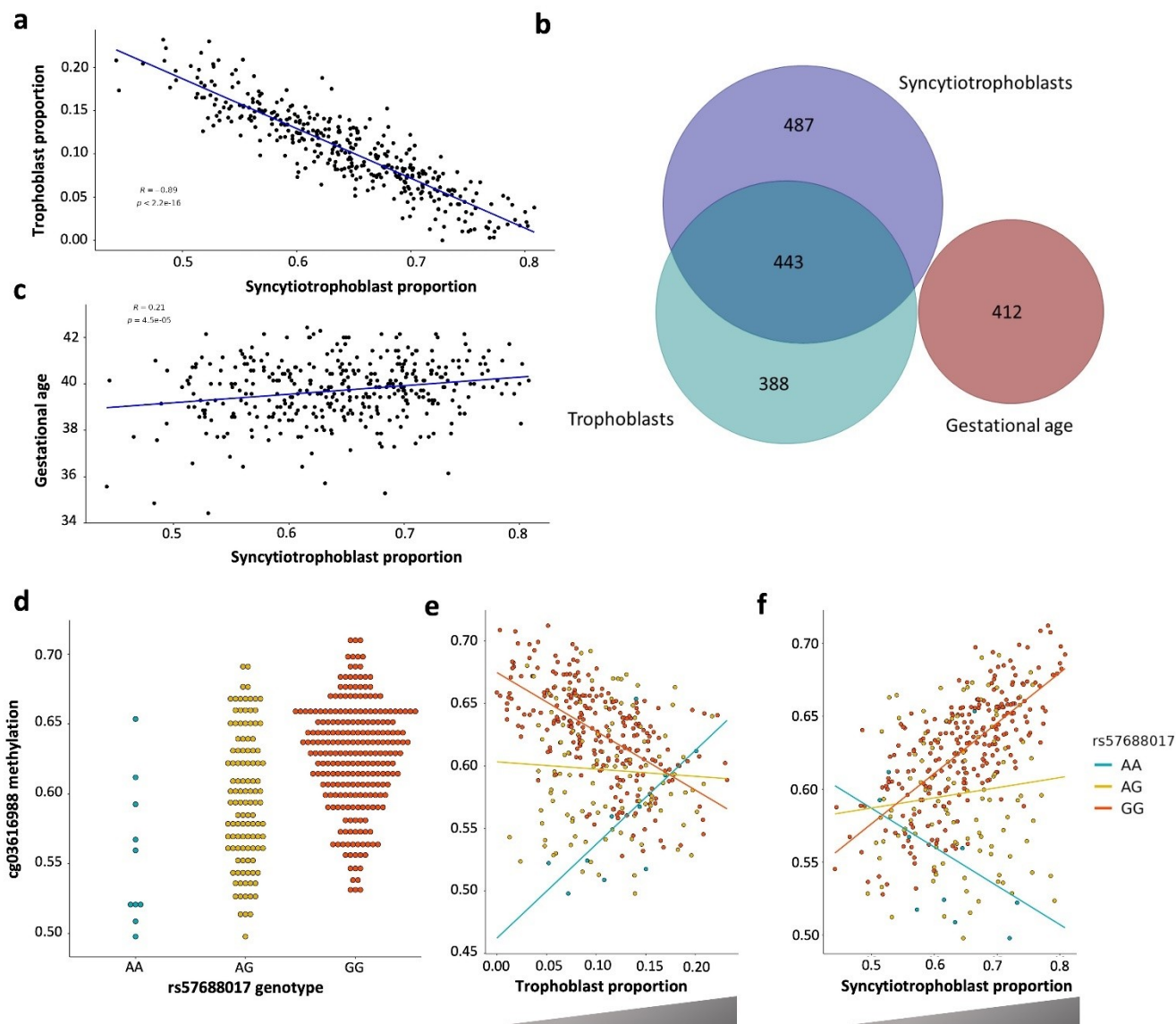
Trait	Abbreviation	N	PubMed ID	Original n° of SNPs
Attention deficit and hyperactivity disorder	ADHD	55,374	30478444	8,094,094
Aggression	AGR	18,988	26087016	2,188,528
Autism spectrum disorder	ASD	46,351	28540026	9,112,386
Bipolar disorder	BIP	413,466	34002096	7,608,183
Internalizing problems	INT	64,561	35378236	5,445,594
Major depression disorder	MDD	500,199	30718901	8,483,301
Obsessive-compulsive disorder	OCD	9,725	28761083	8,409,517
Panic disorder	PD	9,907	31712720	10,151,624
Suicidal attempt	SA	6,183	34737426	11,823,118
Schizophrenia	SCZ	320,404	35396580	7,583,660

## Figures

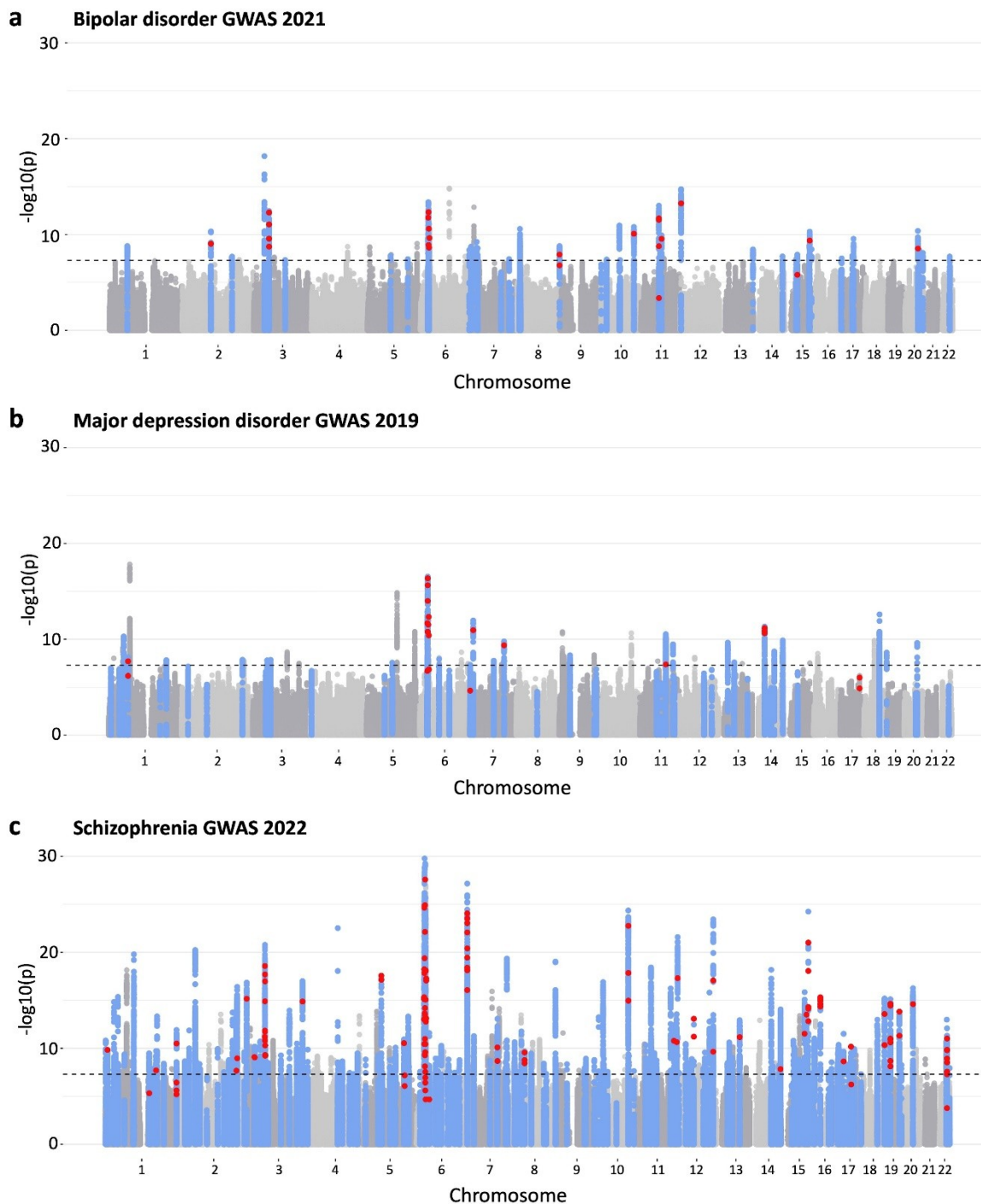


**Figure 1. Characterization of the placental *cis*-mQTLs from the nominal database.** The distance between the SNP-CpG pair participating in the reported *cis*-mQTLs is displayed as a density plot, where the X-axis represents the distance in Mb. The red line represents the median distance of 44 kb.

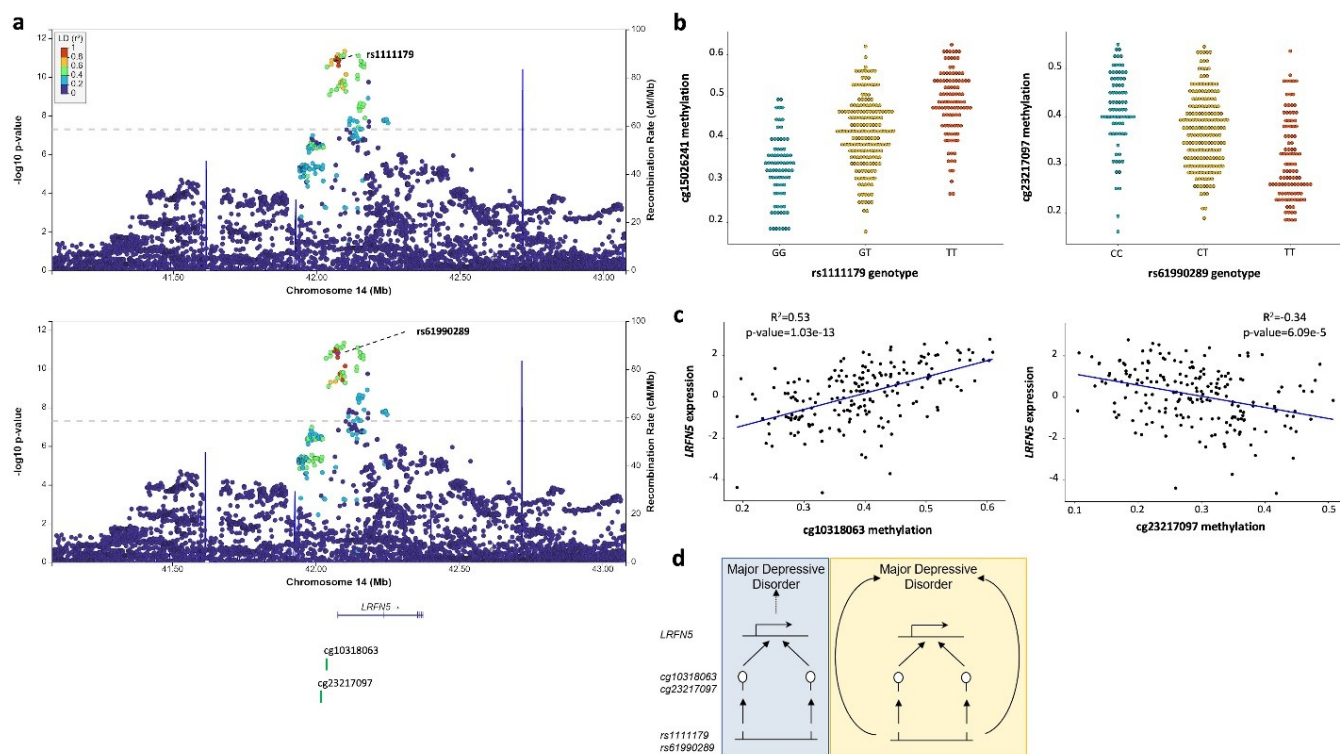
The distribution of the reported *cis*-mQTLs along the chromosomes is shown in the **b** barplot, where the X-axis represents the autosomal chromosomes. The uniform distribution of the effect size from the reported *cis*-mQTLs is pictured in the **c** volcano plot, where the Y-axis illustrates the  $-\log_{10}$  nominal 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 (inner circle) and the nominal mQTL-CpGs (outer circle) considering the Relation To Island and the UCSC RefGene annotation is displayed in the **d** and **e** piecharts, respectively. The methylation beta values, ranging from 0 to 1, of the participating mQTL-CpGs stratified by the Relation To Island annotation is shown in the **f** density plot, where the methylation values are found in the X-axis. The eFORGE enrichment of DNase I hotspots considering the top 10,000 nominal mQTL-CpGs is shown in the **g** plot. The Y-axis represents the  $-\log_{10}$  binomial p-value of the enrichment, and the X-axis to the tissue. False Discovery Rate (FDR) corrected q-values below 0.01 and 0.05 are represented by red and pink dots, respectively, while blue dots show q-values  $>0.05$ .



**Figure 2. STB-, TB- and GA-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 intersection between STB-, TB- and GA-imQTLs is represented in the **b** Venn diagram. The correlation between STB and GA (N=368) is shown in **c**, the X-axis representing the STB proportion in the sample set, and the Y-axis showing the GA. The standard cg03616988-rs57688017 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 cg03616988 DNAm beta values, ranging from 0 to 1. In **d**, the X-axis displays the genotype of rs57688017, while in **e** and **f**, the X-axes show the TB and STB proportion, respectively.

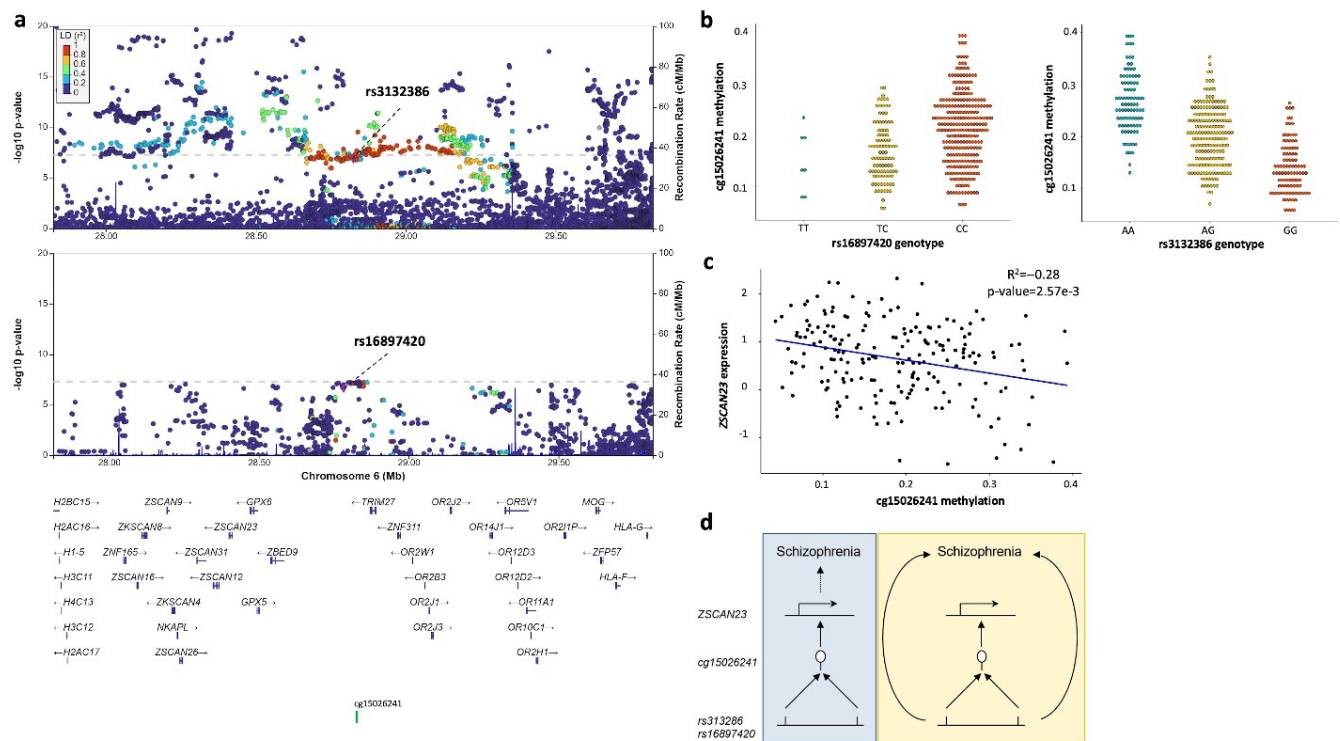


**Figure 3. Manhattan plot of BIP, MDD and SCZ GWAS highlighting the SMR and *coloc* results.** The original GWAS from BIP, MDD and SCZ were plotted in the **a**, **b** and **c** Manhattan plots, respectively. In the Y-axes the original  $-\log_{10}$  p-values are displayed, and in the X-axis the chromosomes. The blue dots represent genomic regions significantly colocalizing with our placental mQTLs, and the red dots are mQTL-SNPs associated with CpGs 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.

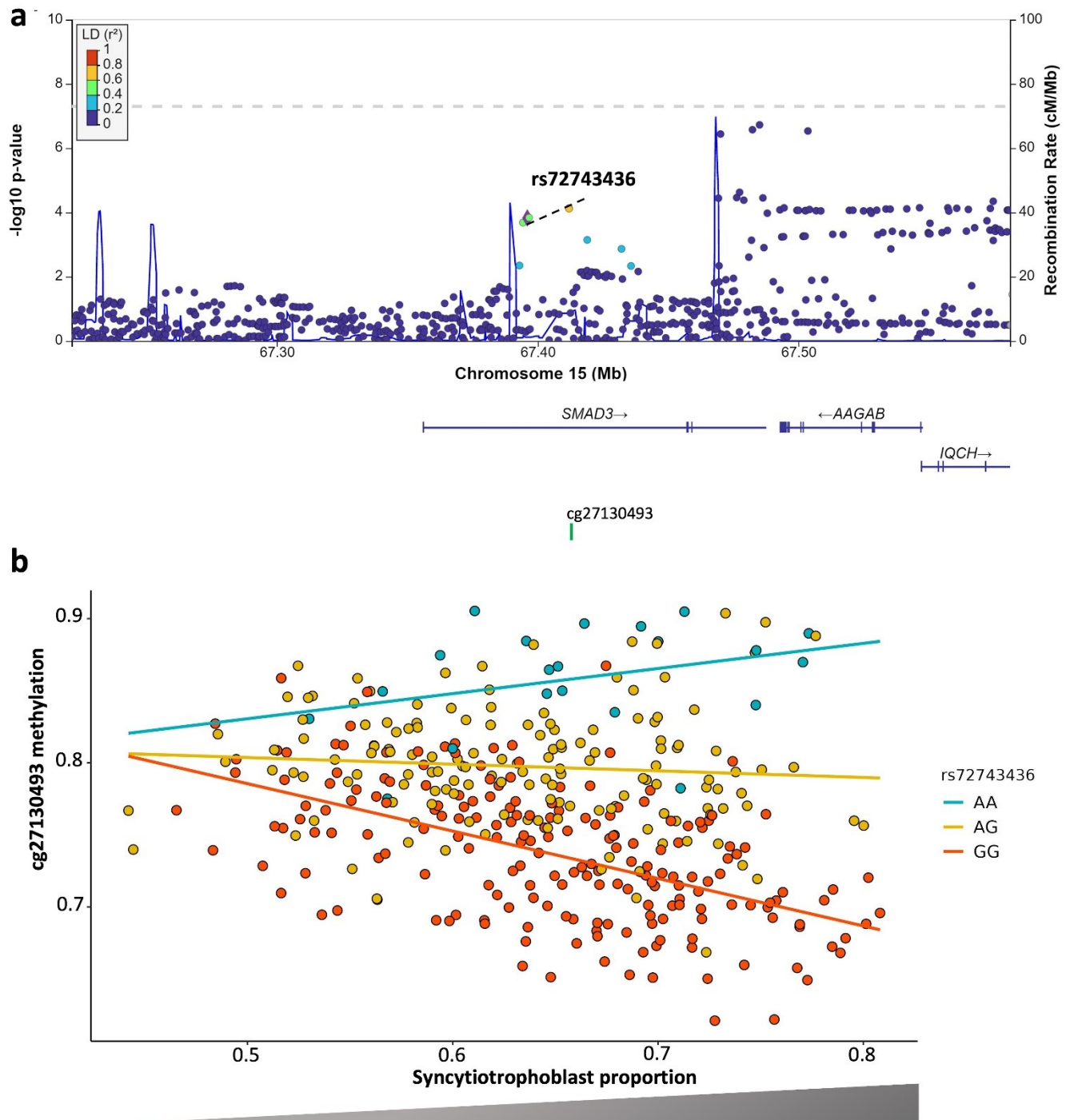


**Figure 4. Two different CpGs pleiotropically associated with MDD.** The mQTL-SNPs rs1111179 and rs61990289, highlighted as purple diamonds, with the  $-\log_{10}$  p-values of the original MDD GWAS are represented in the **a** locusZoom plot. The X-axis displays the involved genomic region of chromosome 14 in Mb, showing the distribution of the coding genes in the locus, as well as the location of the cg10318063 and cg23217097 mQTL-CpGs. The Y-axis shows the  $-\log_{10}$  p-value from the original GWAS, and the SNPs are colour coded as a function of the LD with the highlighted SNP. The two significant mQTLs, rs1111179-cg10318063 and rs61990289-cg32217097, are plotted in the **b** dotplots, where Y-axis represents the beta DNAm values of the indicated CpG, ranging from 0 to 1. The X-axes display the genotype of the indicated SNPs. The eQTM of cg10318063 and cg32217097 mQTL-CpGs are portrayed in the **c** dotplots, where X-axis represents the DNAm values from each involved CpG, ranging from 0 to 1. The Y-axes display the expression values of the *LFRN5* gene in placenta. The hypothesis of the pleiotropical association between the SNPs, the DNAm values of the CpGs in placenta, the gene expression levels of *LFRN5* 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.

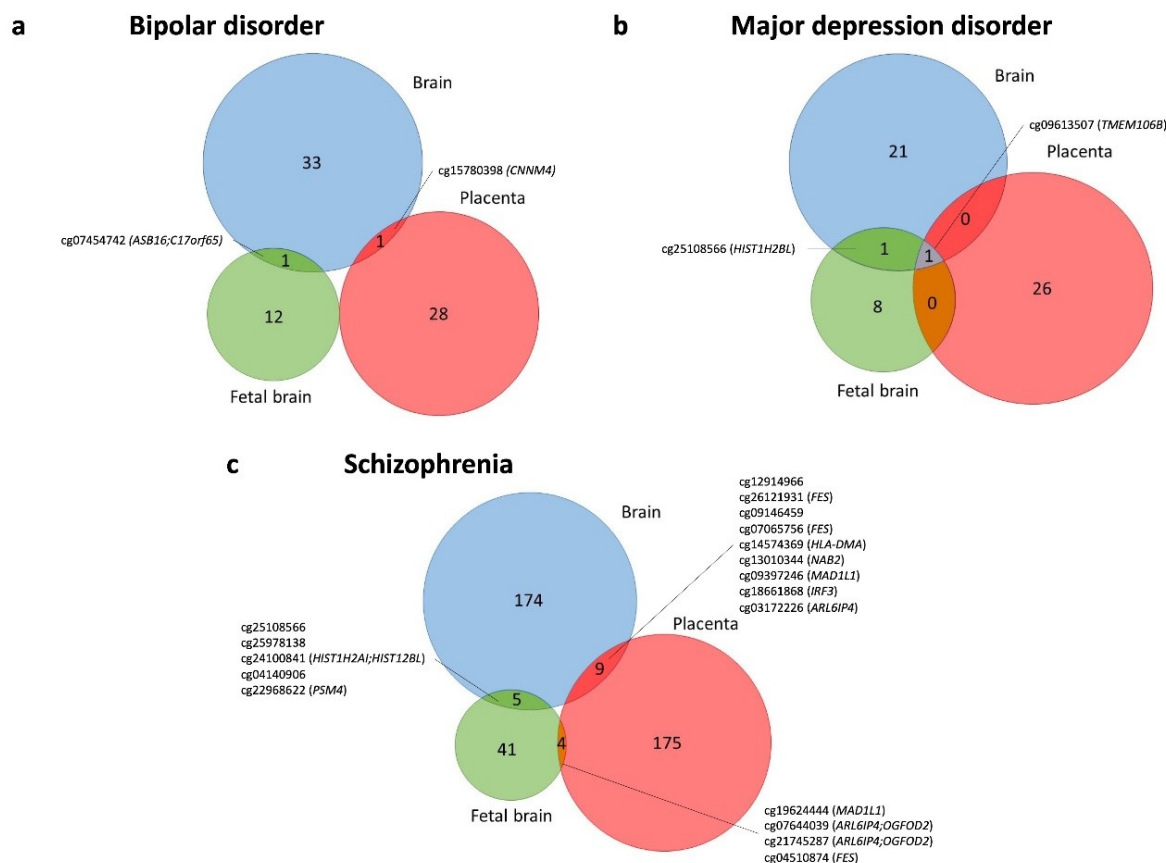




**Figure 5. One CpGs is pleiotropically associated to SCZ, and marked by two independent signals.** The two mQTL-SNPs rs313286 and rs16897420, highlighted as purple diamonds, are shown in the original (top) and the conditional (bottom) SCZ GWAS in the **a** locusZoom plot. The X-axis displays the involved genomic region in chromosome 6 in Mb, showing the distribution of the coding genes in the locus, as well as the location of the mQTL-CpG cg15026241. The Y-axis shows the  $-\log_{10}$  p-value from the original and conditional GWAS, and the SNPs are colour coded as a function of the LD with the highlighted SNP. The two significant mQTLs, rs16897420-cg15026241 and rs3132386-cg15026241, are plotted in the **b** dotplots, where Y-axes represent the cg15026241 beta DNAm values, ranging from 0 to 1. The X-axis displays the genotype of the corresponding SNPs. The eQTM of the significant mQTL-CpG is portrayed in the **c** dotplot, where the X-axis shows the DNAm values of the cg15026241 CpG, ranging from 0 to 1. The Y-axis displays the expression values of the *ZSCAN23* gene in placenta. The hypothesis of the pleiotropical association between the SNPs, the DNAm values of the CpGs in placenta, the gene expression levels of *ZSCAN23* 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.



**Figure 6. STB-imQTL pleiotropically associated with BIP.** The mQTL-SNP rs72743436, 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 mQTL-CpG cg27130493. The Y-axis shows the  $-\log_{10}$  p-value from the original GWAS, and the SNPs are colour coded as a function of LD with the highlighted SNP. The STB-imQTL is pictured in the **b** dotplot. The X-axis represents the cg27130493 beta DNAm values and the Y-axis the STB proportion, both ranging from 0 to 1. The genotype of rs72743436 SNP-imQTL is colour coded as indicated in the legend.



**Figure 7. Overlap between the SMR results in brain, placenta and fetal brain, in BIP, MDD and SCZ.** The overlap between the mQTL-CpGs pleiotropically associated between the three tissues and BIP, MDD and SCZ are represented in the **a**, **b** and **c** Venn diagrams, respectively. Overlapping mQTL-CpGs are shown, and also the closest gene from the Illumina annotation file (between brackets).