

RESEARCH

Open Access



Low reliability of DNA methylation across Illumina Infinium platforms in cord blood: implications for replication studies and meta-analyses of prenatal exposures

Emilie Willoch Olstad^{1,2*}, Hedvig Marie Egeland Nordeng^{1,2,3}, Geir Kjetil Sandve^{2,4}, Robert Lyle^{2,5,6} and Kristina Gervin^{1,2,7}

Abstract

Background: There is an increasing interest in the role of epigenetics in epidemiology, but the emerging research field faces several critical biological and technical challenges. In particular, recent studies have shown poor correlation of measured DNA methylation (DNAm) levels within and across Illumina Infinium platforms in various tissues. In this study, we have investigated concordance between 450 k and EPIC Infinium platforms in cord blood. We could not replicate our previous findings on the association of prenatal paracetamol exposure with cord blood DNAm, which prompted an investigation of cross-platform DNAm differences.

Results: This study is based on two DNAm data sets from cord blood samples selected from the Norwegian Mother, Father and Child Cohort Study (MoBa). DNAm of one data set was measured using the 450 k platform and the other data set was measured using the EPIC platform. Initial analyses of the EPIC data could not replicate any of our previous significant findings in the 450 k data on associations between prenatal paracetamol exposure and cord blood DNAm. A subset of the samples ($n = 17$) was included in both data sets, which enabled analyses of technical sources potentially contributing to the negative replication. Analyses of these 17 samples with repeated measurements revealed high per-sample correlations ($\bar{r}_R \approx 0.99$), but low per-CpG correlations ($\bar{r}_R \approx 0.24$) between the platforms. 1.7% of the CpGs exhibited a mean DNAm difference across platforms > 0.1 . Furthermore, only 26.7% of the CpGs exhibited a moderate or better cross-platform reliability (intra-class correlation coefficient ≥ 0.5).

Conclusion: The observations of low cross-platform probe correlation and reliability corroborate previous reports in other tissues. Our study cannot determine the origin of the differences between platforms. Nevertheless, it emulates the setting in studies using data from multiple Infinium platforms, often analysed several years apart. Therefore, the findings may have important implications for future epigenome-wide association studies (EWASs), in replication, meta-analyses and longitudinal studies. Cognisance and transparency of the challenges related to cross-platform studies may enhance the interpretation, replicability and validity of EWAS results both in cord blood and other tissues, ultimately improving the clinical relevance of epigenetic epidemiology.

*Correspondence: e.w.olstad@farmasi.uio.no

¹ Pharmacoepidemiology and Drug Safety Research Group, Department of Pharmacy, Faculty of Mathematics and Natural Sciences, University of Oslo, Oslo, Norway

Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Keywords: Epigenetic epidemiology, Epigenetics, EWAS, MoBa, MBRN, Validity, Replication, Reliability, Illumina Infinium platforms, Microarrays

Background

Epigenetics entails modifications of the DNA that can impact gene expression, but does not involve changes in the underlying DNA sequence. The most commonly studied epigenetic modification is DNA methylation (DNAm), which occurs at cytosine bases of cytosine–phosphate–guanine dinucleotide sites (CpGs). DNAm can be impacted by the DNA sequence, as well as environmental influences [1–4]. There is an increasing interest in the role of epigenetics within epidemiology. Several pharmacoepidemiological studies have reported an association between prenatal psychotropic or analgesic medication exposure and neurodevelopmental outcomes in the offspring [5–13]. Furthermore, multiple epigenome-wide association studies (EWASs) have identified DNAm changes associated with medication exposure during pregnancy (e.g. valproic acid, antidepressants and paracetamol) [14–20]. Recently, we found an association between cord blood DNAm and prenatal long-term exposure to paracetamol in children with attention-deficit/hyperactivity disorder (ADHD) [21]. These initial findings may suggest that DNAm is involved in the relationship between prenatal medication exposure and adverse neurodevelopmental outcomes [3, 4].

Despite a growing interest in epigenetics, and an increasing number of published EWASs, there are several critical biological and technical challenges in epigenetic epidemiology, which have important implications for the interpretation, validity and clinical translation of the findings [1, 22, 23]. One key challenge is the paucity in the replication of findings. For instance, two systematic literature reviews on the association of offspring epigenetic patterns with medication use [20] and maternal well-being in pregnancy [24] uncovered largely inconsistent findings. These reviews suggest multiple origins of the discrepant results, such as small sample sizes resulting in low statistical power and poor study designs [20, 24]. The majority of EWASs are based on DNAm data generated using the Illumina Infinium HumanMethylation Bead-Chip platforms, including the 27 k ($n > 27,000$ CpGs), 450 k ($n > 450,000$ CpGs) and the EPIC arrays ($n > 850,000$ CpGs) [25]. Recent studies have elucidated technical aspects related to the Infinium platforms, which have significant influences on the analyses and interpretation of results. These studies have shown significant per-CpG differences and poor per-CpG correlation both within [26–35] and across [31, 32, 36–40] microarray platforms, which challenges combined analyses of DNAm data from

both platforms (e.g. [41–45]). In cord blood, the median correlation of individual CpGs across platforms was only 0.24 [37]. Furthermore, 2.4% of the CpGs exhibited a mean difference in measured DNAm level between the platforms ≥ 0.1 [37], on the same order as the low effect sizes often observed within epigenetic epidemiology [1, 22, 46]. Furthermore, only 18.0% of CpGs in adult whole-blood exhibit a moderate or better reliability across platforms (intra-class correlation coefficient [ICC] ≥ 0.5) [31]. The technical aspects contributing to low reliabilities of CpGs may affect the power of EWASs [28, 47]. Consequently, poor concordance of measured DNAm levels across platforms may impact both the replicability and validity of EWAS results.

In an ongoing study, we aim to replicate and expand our previous findings showing associations between long-term prenatal exposure to paracetamol (≥ 20 days) and DNAm in children with ADHD [21]. Analyses of DNAm data generated from a larger number of samples selected from the same cohort using the Infinium EPIC platform find no significant CpGs associated with paracetamol exposure. Accordingly, we fail to replicate any of our previous significant findings [21]. Examining a subset of samples with repeated measurements in both data sets has enabled a thorough investigation of potential technical origins of the negative replication. Our findings could not explain the failure to replicate our previous results, but are still important for replication EWASs, as well as studies combining DNAm from different Infinium platforms, such as longitudinal studies or meta-analyses.

Results

Lack of replicability may originate from several technical sources

This study is based on a subset of samples ($n = 17$) included in two datasets and consists of repeated measurements using the Infinium 450 k and EPIC platforms. The samples were selected from the Norwegian Mother, Father and Child Cohort Study (MoBa). In the data set assessed on the 450 k platform ($n = 384$ samples), we have previously published associations between prenatal exposure to paracetamol and DNAm differences in children with ADHD [21]. Analysis of the second data set ($n = 261$ samples), which was designed to expand on these findings using the EPIC platform, has failed to replicate our previous findings (data not shown). This prompted a thorough investigation of whether technical aspects of the Infinium platforms could explain the negative replication. Using

a subset of samples with repeated measurements from both studies ($n=17$ samples), we conducted systematic analyses to assess the integrity and reliability of the DNAm data between the Infinium platforms.

The DNAm data separate into clusters explained by microarray platforms

We performed stringent quality control, normalisation and probe filtering procedures of the DNAm data from the two data sets containing the samples with repeated measurements, to minimise technical variation related to pre-processing of the data. First, we examined DNAm data measured for a set of genotyping probes on each platform ($n=59$ probes). Clustered heatmaps of DNAm values at these genotyping probes showed that the repeated cross-platform measurements of each sample grouped together and hence, excluded potential mix-up of samples (Additional file 1: Fig. S1). Second, we examined whether pre-processing steps such as background and probe-type correction impacted the cross-platform concordance. To do this, we used the intra-class correlation coefficient (ICC), which equals 1 if there is perfect per-CpG concordance between the measured DNAm in the 450 k and EPIC data sets. Generally, an $ICC < 0.5$ is considered poor [48, 49]. We computed the ICCs after

pre-processing the 450 k and EPIC data sets separately, using the default settings of five commonly used pre-processing pipelines *ChAMP* [50, 51], *ENmix* [34], *minfi* [52], *RnBeads* [53] and *wateRmelon* [54] (Additional File 1: Table S1). We also included one pipeline commonly reported in the literature, namely *RnBeads* with the background and probe-type corrections *ENmix.oob* [34] and *BMIQ* [55], respectively. This analysis revealed that the *ENmix* pipeline exhibited larger ICCs than the other pipelines (Fig. 1). Therefore, we performed the rest of the analyses on data sets normalised using the default settings of the *ENmix* pipeline.

Next, we performed principal component analysis (PCA) to explore technical variation in the DNAm data related to the 450 k and EPIC platforms. As expected, PCA revealed distinct clustering of samples corresponding to the 450 k and EPIC platforms (Fig. 2). Similar plots were observed when pooling all the available 450 k and EPIC samples ($n=628$ samples; data not shown).

DNAm levels differ between the 450 k and EPIC platforms

To further investigate the dissimilarities between the 450 k and EPIC platforms, we computed the difference in and correlation of DNAm at overlapping CpGs on the two platforms ($n=397,813$ CpGs). These analyses

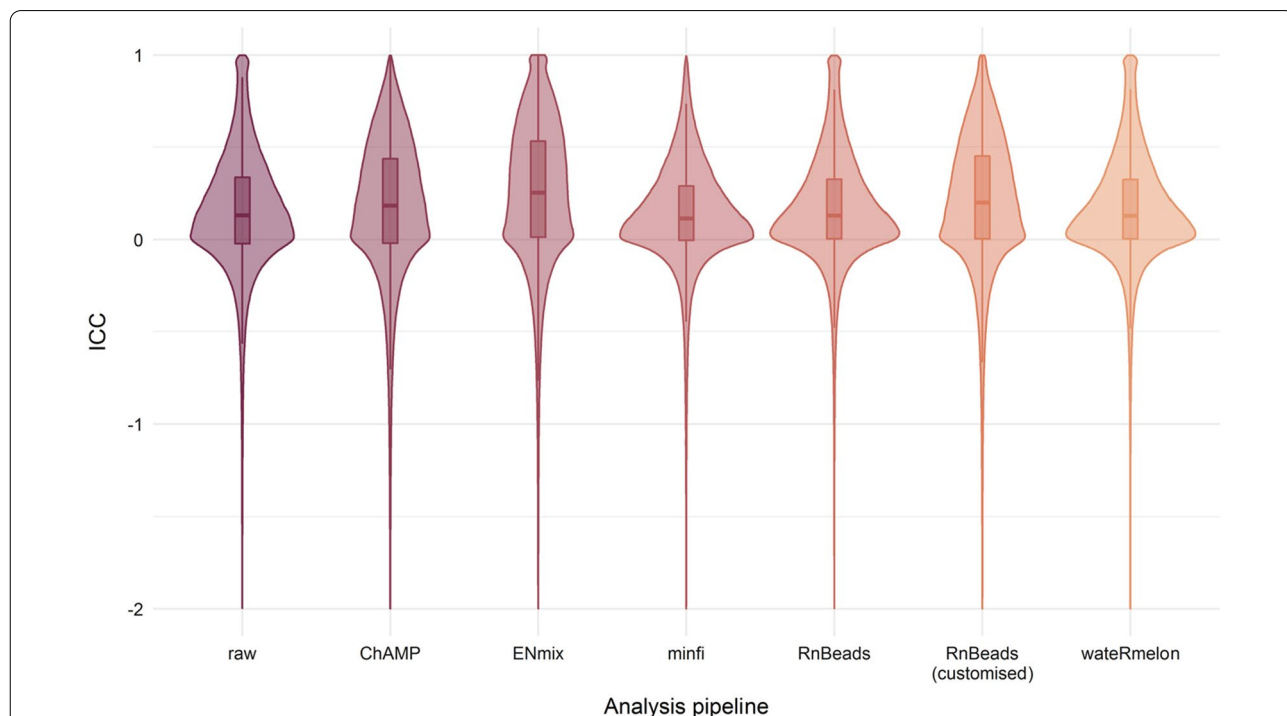
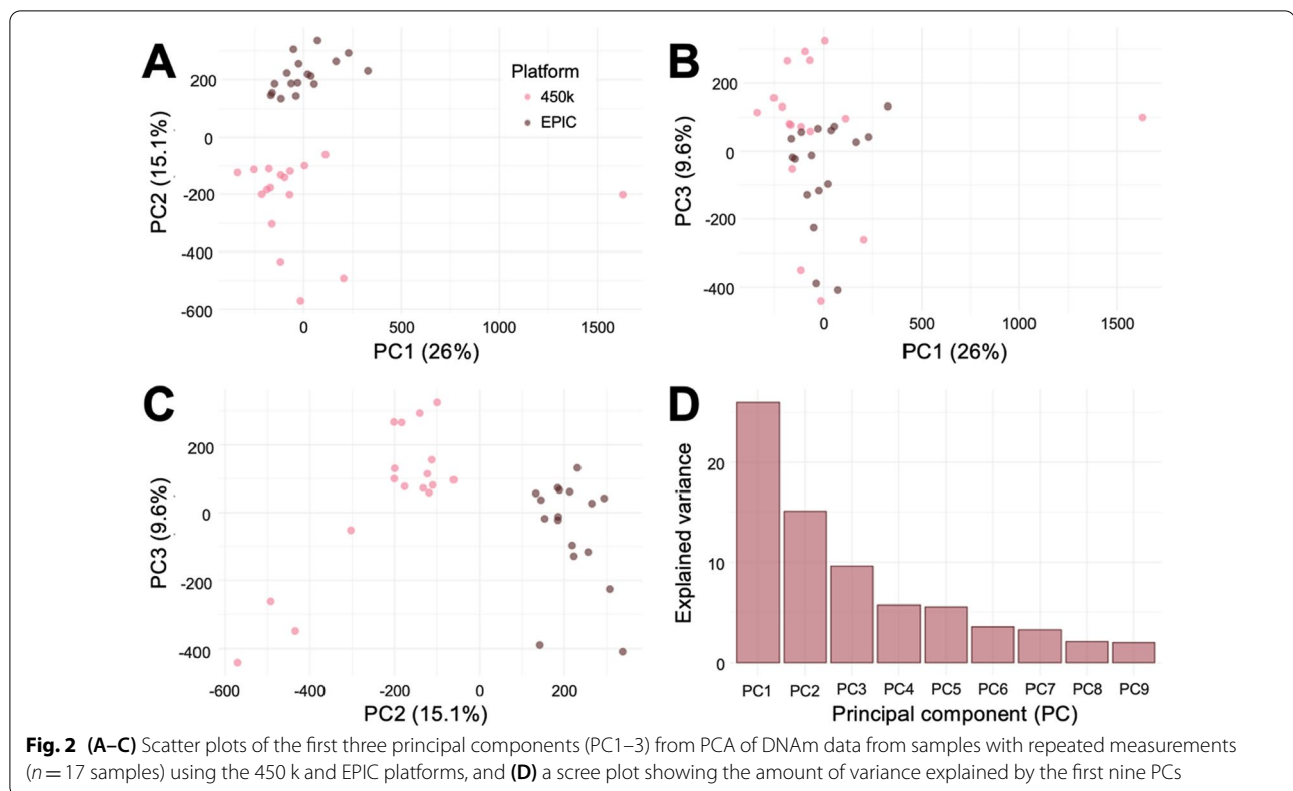


Fig. 1 Overview of the ICC distribution computed from raw data and from data pre-processed using the default settings of five common EWAS analysis pipelines. Additionally, we included one common analysis pipeline (“RnBeads (customised)”, including the normalisation methods *ENmix.oob* and *BMIQ*). All pipelines examined also exhibited ICCs lower than -2 , but these were removed from the illustration for visualisation purposes. The default settings of each analysis pipeline are detailed in Additional file 1: Table S1



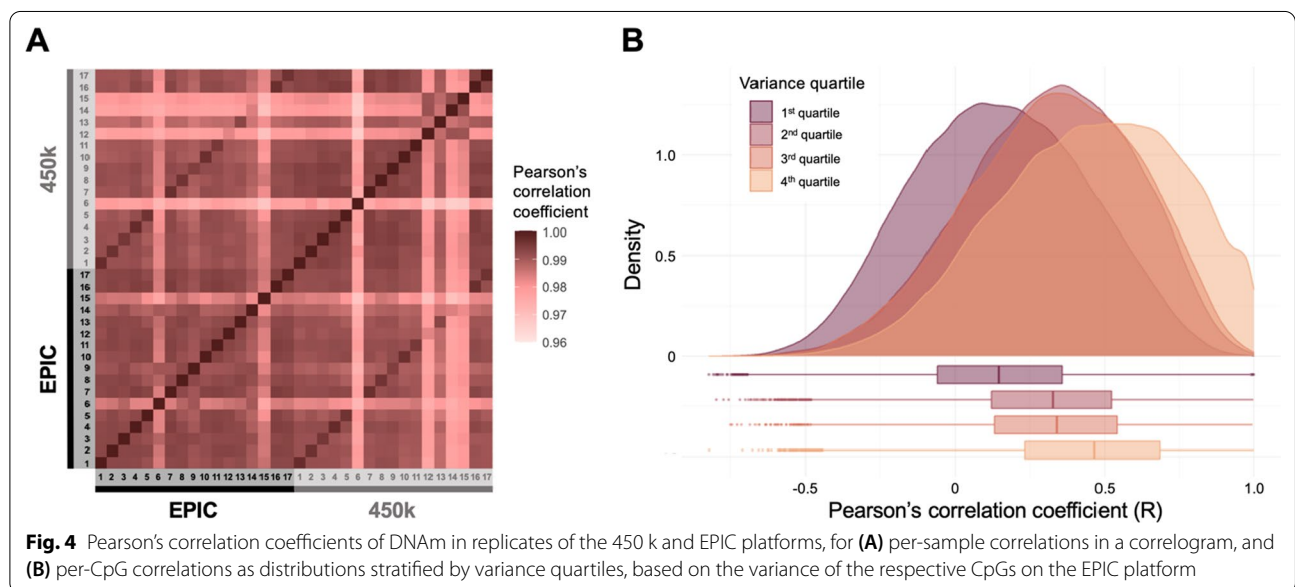
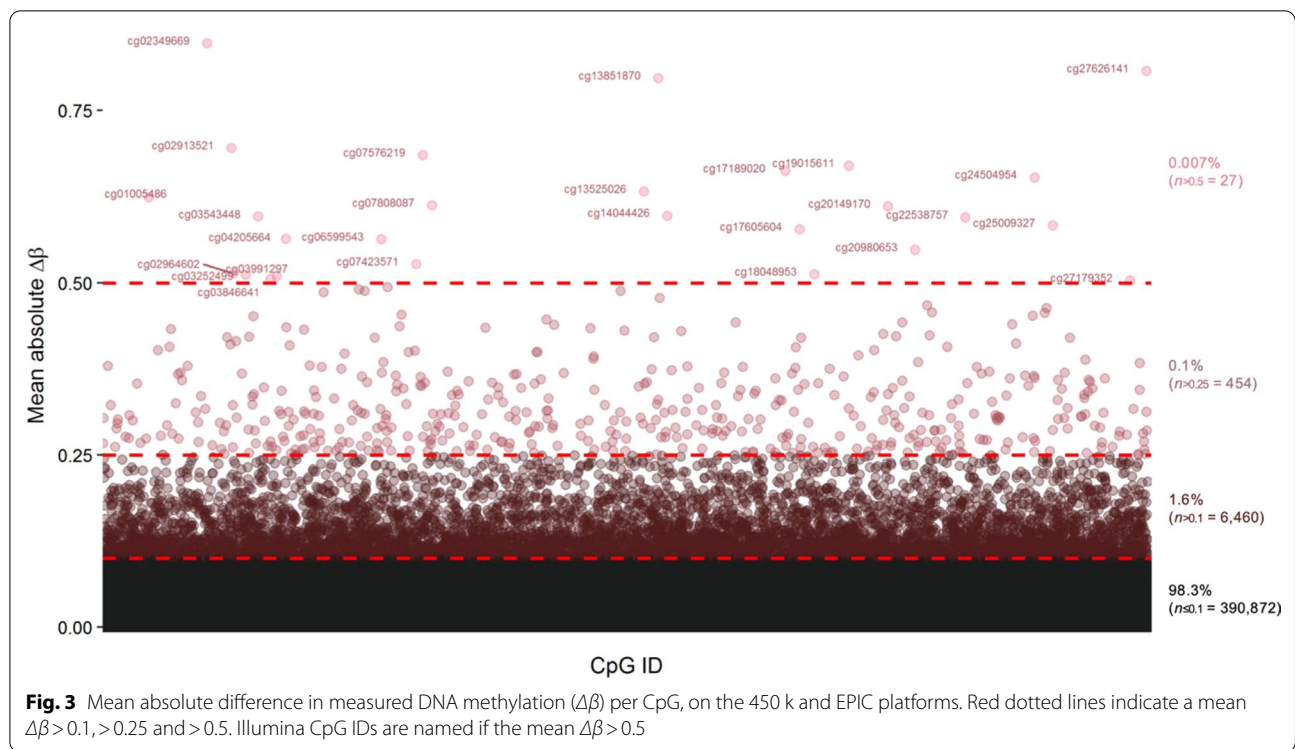
revealed small per-sample absolute differences in DNAm at overlapping CpGs between the two arrays (median ≈ 0.008 and mean ≈ 0.017 absolute DNAm differences). For 0.1% ($n = 454$) of CpGs, the mean DNAm difference over all replicates was > 0.25 , while 0.007% ($n = 27$) of CpGs exhibited a mean DNAm difference > 0.5 (Fig. 3). These numbers are largely in line with previous studies, comparing differences in measured DNAm between the 450 k and EPIC arrays in cord blood [37], whole-blood [31, 32, 36, 37], placenta [38] and cartilage [39]. Furthermore, of the 27 CpGs with an absolute mean DNAm difference > 0.5 , 5 of these CpGs also exhibited absolute mean DNAm difference > 0.5 in both cord blood [37], whole-blood [37], placenta [38] and cartilage [39] (Additional file 1: Fig. S2).

We observed a high per-sample correlation of DNAm between the platforms, both when comparing replicates, and when comparing two independent samples across the platforms (Fig. 4A). The median per-sample Pearson's correlation coefficient was 0.996 and the mean was 0.992, with the lowest correlation between any two samples being 0.969 and the highest being 0.998. In contrast, the per-CpG correlations of measured DNAm between the platforms were significantly lower: the median correlation was 0.237 and the mean was 0.238, with the lowest correlation being -0.822 and the highest being 1.00

(Fig. 4B). The per-CpG correlation appeared to be related to the variance of each CpG, which were similar for both platforms; CpGs with high correlation also exhibited larger variance (Fig. 4B). The high per-sample correlation, low per-CpG correlation, and the relationship between CpG variance and correlation, have previously been reported for cord blood [37], and multiple other tissues [31, 32, 36–39].

Few CpGs are reliable between the 450 k and EPIC platforms

In order to examine concordance of cross-platform DNAm levels, we assessed the reliability of the CpGs, reflecting both correlation and agreement. To do this, we computed the ICC, as previously suggested by Sugden et al. (2020) comparing cross-platform DNAm levels in adult whole-blood [31]. Overall, the ICCs of the overlapping CpGs were poor (median = 0.246 and mean = 0.230; Fig. 5A). Approximately 26.7% ($n = 106,078$) of the CpGs exhibited an $ICC \geq 0.5$. This is similar to the findings of the recent study by Sugden et al. in adult whole-blood, where 18.0% of CpGs exhibited an $ICC \geq 0.5$ [31]. Approximately 38.6% ($n = 40,916$) of the CpGs with an $ICC \geq 0.5$ in the current study overlapped with the CpGs with an $ICC \geq 0.5$ reported by Sugden et al. [31] (Additional File 2). The microarray type II probes exhibited slightly better ICCs and correlation coefficients than type



I probes (Additional File 1: Fig. S3). Probes with poor ICCs and correlation coefficients appear more frequently in CpG islands (Additional File 1: Figs. S4 and S5), possibly due to an increased proportion of largely unmethylated CpGs in these regions (Additional File 1: Fig. S6).

Considering the poor CpG reliabilities, we investigated if the ICCs of the repeated measurements were

higher than expected for two randomly paired samples. Therefore, we paired each EPIC sample with a randomly selected 450 k sample. The distribution of ICCs computed from the 17 repeated measurements (Fig. 5A) is significantly different from the ICC distributions computed from the 17 random 450 k-EPIC pairs (Kolmogorov–Smirnov test: $p < 2.2 \cdot 10^{-16}$; Additional file 1: Fig. S7).

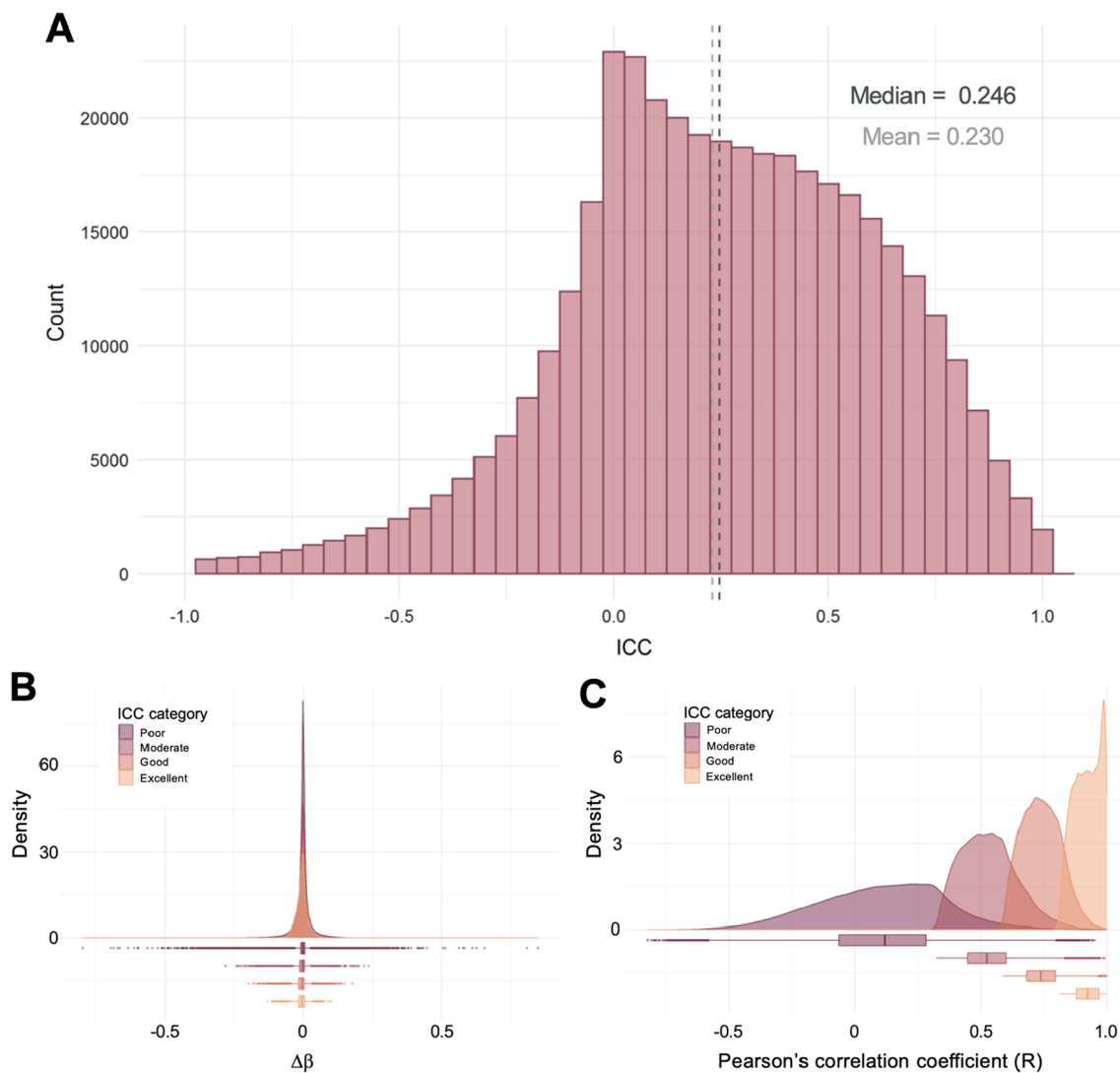


Fig. 5 (A) Histogram of the ICCs computed from the 17 samples assessed on both the 450 k and EPIC platforms. (B) Density distribution of mean difference in DNAm level, stratified by ICC category. (C) Density distribution of Pearson's correlation coefficient, stratified by ICC category. The ICC categories are defined as follows: poor: $ICC < 0.5$; moderate: $0.5 \leq ICC < 0.75$; good: $0.75 \leq ICC < 0.9$; and excellent: $ICC \geq 0.9$. The dark grey, dotted line indicates the median ICC, and the light grey, dotted line indicates the mean ICC. Outlying CpGs with ICCs less than the mean ICC minus three standard deviations were removed for visualisation purposes, but were included for summary statistic calculations

Furthermore, only a small percentage of the CpGs of the random pairs (2.4%–4.8%) exhibited an $ICC \geq 0.5$, which are significantly different proportions from the ICCs of the repeated measurements (Pearson's Chi-squared test: $p < 2.2 \times 10^{-16}$).

The ICC reflects both correlation and agreement across microarray platforms

To investigate if the ICCs reflect both agreement and correlation across platforms, we examined the distribution of mean differences in DNAm and Pearson's correlation coefficients, for each of four ICC categories:

poor ($ICC < 0.5$), moderate ($0.5 \leq ICC < 0.75$), good ($0.75 \leq ICC < 0.9$) and excellent ($ICC \geq 0.9$) [48]. The distribution of mean differences in DNAm is relatively similar between the ICC categories. However, there are far more of the poor CpGs displaying large differences in mean DNAm levels across platforms compared to the other ICC categories (Fig. 5B). In contrast, the correlation coefficient increases with improving ICC category; the poor ICC category exhibits a wide range of low correlation coefficients (median ≈ 0.12), while the distribution of the correlation in the excellent category is highly skewed to the right (median ≈ 0.92). The moderate and

good categories exhibit a wider range of correlation coefficients than the excellent CpGs, with a median of 0.52 and 0.74, respectively (Fig. 5C).

These observations demonstrate that the reliability of each CpG depends on both the correlation and the agreement between the two platforms [48]. An excellent CpG will have both a low mean difference in DNAm between platforms and a high correlation, explaining the small range in values of both the mean DNAm differences and the correlation coefficients. In contrast, a poor probe (including a larger range of ICCs) may exhibit an acceptable correlation but have a large mean DNAm difference (Additional file 1: Fig. S8). For instance, 685 of the 5407 CpGs with an $R \geq 0.9$ nevertheless have an $ICC \leq 0.9$, with 22 CpGs even having a poor $ICC (< 0.5)$. Furthermore, of the 395,286 CpGs with a mean DNAm difference ≤ 0.1 , 289,327 exhibit a poor $ICC (< 0.5)$. This is likely due to low correlations, as the median R for these CpGs is 0.12, while the median R was 0.59 for the 105,959 CpGs with a mean DNAm difference ≤ 0.1 and an $ICC \geq 0.5$. Hence, the ICC better reflects reliability across platforms than either accuracy or correlation on their own.

The significant CpGs in the 450 k data have low reliabilities

We then asked if our failure to replicate the findings in our original study [21] could be explained by poor-performing probes, by examining the ICCs of the significant CpGs from the 450 k data set. The significant CpGs for the three group comparisons performed in the original study have median ICCs of 0.119, 0.122 and 0.135 (Additional file 1: Fig. S9). These reliabilities are low compared to the overall mean and median of the ICCs including all common CpGs across platforms.

Discussion

Replication of association studies is important to ensure robust and valid findings. In an ongoing study, we aimed to replicate and expand on findings in our previous study, where we found an association between long-term prenatal paracetamol exposure and differences in DNAm in children with ADHD, using the Infinium 450 k platform [21]. Surprisingly, analyses of the follow-up data consisting of a larger sample and use of the Infinium EPIC platform have not replicated the results from our original study. Indeed, a challenge of EWASs is to discern spurious findings from true positives, rendering the replication of significant associations challenging [1, 22, 23]. Recent studies have shown low concordance across 450 k and EPIC platforms in different tissues [31, 32, 36–40]. Therefore, we have conducted a systematic evaluation of technical aspects related to concordance of DNAm data across the Infinium platforms in our studies

in cord blood, by using data from a subset of samples with repeated measurements from the 450 k and EPIC platforms.

Technical variation such as batch effects is systematic variation caused by, for example, processing by different technicians, varying reagent batches and differences in the scanner performance. PCA of DNAm data from the samples with repeated measurements demonstrated distinct clustering of samples corresponding to the platform. If these differences in DNAm were independent of the platform and resulted entirely from positioning on the beadchip or bisulphite conversion plate, we would expect the changes to be relative and to not impact the replicability. Considering the general challenge of replication of EWASs [1, 22, 23] and the low per-CpG concordance across platforms reported in several recent studies [31, 32, 36–40], we were encouraged to examine possible cross-platform differences in DNAm. Corroborating previous studies, we observed a high per-sample correlation even between the randomly paired samples [32, 36–40]. In contrast, the per-CpG correlation was significantly lower, and some probes exhibited large differences in mean measured DNAm for overlapping CpGs on the two platforms.

Considering the highly concerning findings by Sugden et al. [31], reporting low reliabilities (measured by ICCs) for most CpGs across the 450 k and EPIC platforms in adult whole-blood, we estimated the ICCs of each CpG across the two platforms in our cord blood samples. Ideally, the ICC will approach 1 if the between-sample variation is much larger than the within-sample variation, suggesting larger biological variation than technical variation. However, most CpGs in our study exhibited poor reliabilities ($ICC < 0.5$) [31, 48] and we found that only 26.7% of CpGs in cord blood had an acceptable reliability across platforms. Interestingly, 38.6% of these CpGs overlapped with the 18.0% reliable CpGs identified in adult whole-blood [31]. This may suggest that some probes are generally unreliable in different tissues, possibly due to cell-type specific variability in DNAm. In contrast, other CpGs may perform worse in specific tissues, similar to what has been suggested for both per-CpG correlations and differences in DNAm between platforms [37–39]. In future studies, it would be interesting to examine the ICCs between Infinium platforms and other DNAm measuring technologies, such as whole-genome bisulphite sequencing (WGBS) or methylated immunoprecipitation (MeDIP).

We observed a substantial difference in the distribution of ICCs for different pre-processing steps used in common analysis pipelines. The *ENmix* pipeline exhibited the largest median ICC , suggesting that this pipeline may be better to best conserve the similarity of normalised

repeated measurements from different platforms. In contrast, both the default *RnBeads*, *minfi* and *wateRmelon* pipelines have no better ICC distributions than the raw data. Notably, compared to a recent study reporting the ICC distribution of multiple different pipelines for within-platform repeated measurements [35], the distribution of cross-platform ICCs varies more dependent on the analysis pipeline used. However, the analysis pipeline with the highest median ICC is *ENmix* for both cross-platform and within-platform comparisons [35].

Interestingly, some studies have reported that cross-platform differences in DNAm and poor per-CpG correlations do not substantially impact the outcome of EWASs [32, 37]. However, when investigating the relationship of ICCs with the likelihood of replication of CpGs, Sugden et al. observed a positive relationship between increasing ICC and increasing replication rate for the association of DNAm with smoking [31]. Similar associations of ICCs with replicability have been found when ICCs were estimated from 450 k-450 k replicates [26, 49]. For instance, smoking-DNAm associations in whole-blood are highly replicable [56], and in one study, 96% of CpGs associated with smoking exhibit high reliability [26]. Additionally, poor ICCs have been shown to decrease the power of individual CpGs in EWASs, i.e. reducing the positive predictive value (PPV) by decreasing the number of true positives [28, 31, 47]. The median ICC of the significant CpGs in our original study was poor. However, if these findings were explained by the low reliability of the probes, we would expect none or very few significant CpGs. Consequently, based on the calculated ICCs using our 17 samples with repeated measurements, we have no explanation for the lack of replicability of our original findings.

A limitation of the present study is the small sample size used to assess the ICCs. However, ICC calculations generally require relatively small sample sizes [47, 57], and Sugden et al. found that sample sizes as small as 25 would be sufficient to detect 80% of all CpGs with an $ICC \geq 0.75$ [31]. Furthermore, our results on both per-CpG correlations, differences in mean DNAm and ICCs are in line with other studies reporting one or more of these measurements for various tissues [31, 32, 36–40]. Nevertheless, a study including a larger number of repeated measurements in cord blood across the 450 k and EPIC platforms should be performed to strengthen our findings. Another limitation of our study is our inability to assess which technical variable(s) associated with the platform are contributing to the differences between platforms. Firstly, the DNAm on the 450 k and EPIC platforms was measured three years apart. Yet, this largely emulates the setting in most studies relying on data processed at different times and in different facilities (e.g.,

longitudinal studies and meta-analyses). Furthermore, all samples included in the current study were processed in the same core facility and by the same technician. Secondly, batches of bisulphite conversion reagents and scanners may also contribute to the cross-platform differences. Nevertheless, we expect that such technical variation is relative within the platforms and, consequently, that probes are mainly affected equally within the platform. Finally, it is challenging to assess the potential contribution of sample plate and beadchip to cross-platform differences, due to the different platform layouts (the 450 k beadchip can load 12 samples, while the EPIC beadchip can load 8 samples). To limit the contribution of variation from sample plate and beadchip in our data, the samples were randomly positioned on plates and beadchips. Accordingly, technical variation contributed by these variables should be random and should not inflict much bias when comparing DNAm between platforms.

The substantial differences across platforms revealed in this and previous studies [31, 32, 36–40] are troubling when trying to replicate findings using a different platform than in the original study. Replication of findings has long been considered a major challenge within epigenetic epidemiology [1, 22, 23], and to our knowledge, only one study has highlighted the potential impact of unreliable CpGs for replication of findings using data from different microarray platforms [31]. Challenges associated with differences in mean DNAm levels across platforms are not necessarily limited to issues of replication. For instance, longitudinal studies based on DNAm measured at multiple timepoints may suffer under the development of new microarray technologies (e.g., [41, 42]). Furthermore, this is also relevant for large meta-analyses combining data from multiple cohorts to increase the power of EWASs (e.g. [43, 44]), often based on large consortia such as the Pregnancy And Childhood Epigenetics (PACE) consortium [45]. Such strategies may be impacted by unreliable probes when combining data sets from different platforms. Similarly, unreliable CpGs across platforms may have implications for current EWAS knowledgebases, such as the EWAS Atlas [58] and the EWAS catalogue [59], which curate EWAS publications to report DNAm-trait associations.

Conclusion

In conclusion, our failure to replicate significant CpGs associated with prenatal paracetamol exposure prompted a thorough investigation of potential technical origins of our null findings. The observation of low cross-platform per-CpG correlation and reliability corroborates previous reports. However, the low-reliability probes could not explain the inability to replicate previous findings in our case. Nevertheless, the poor cross-platform reliabilities

may have important implications for future EWASs, in replication, meta-analyses and longitudinal studies. Therefore, we encourage researchers performing EWASs to examine the reliability of probes within and across tissues and to establish which probes are most comparable across microarray platforms. However, in many cases, the availability of repeated measurements from individual samples may be limited for reasons such as extra cost and limited availability of sample material. To this end, we encourage joint efforts to more extensively outline reliable probes in different tissues. If such investigations reveal common poor-performing probes across or within tissues, other studies may rely on this information when performing cross-platform studies. We hope our findings, supporting the results by Sugden et al. [31], increase awareness of possible challenges in including both 450 k and EPIC data in the same study. Cognisance and transparency of these challenges as well as appropriate precautions when performing cross-platform epigenetic investigations, may enhance the interpretation, replicability and validity of results, and could ultimately improve the clinical relevance of epigenetic epidemiology.

Methods

Sample population

We analysed cord blood samples from the Mother, Father and Child Cohort Study (MoBa). MoBa is a population-based pregnancy cohort study conducted by the Norwegian Institute of Public Health (NIPH) [60–63]. Participants were recruited from all over Norway from 1999–2008 [60, 61]. The women consented to participation in 41% of the pregnancies [60, 61]. The cohort includes approximately 114,500 children, 95,200 mothers and 75,200 fathers [60, 61]. The current study is based on Data Version 8 of the quality-assured data files released for research in 2015. Observational data from MoBa questionnaires Q1 (gestational week 0–13), Q3 (gestational week 13–29) and Q4 (gestational week 30 to delivery) were used to select individuals for the study. The personal, 11-digit identification number, unique to every permanent resident of Norway, was used to link MoBa to the Norwegian Patient Registry (NPR) and the Medical Birth Registry of Norway (MBRN). MBRN is a national health registry containing information about all births in Norway. We also analysed umbilical cord blood samples retrieved from the MoBa biobank [62, 63]. The biobank stores blood samples obtained from both parents during pregnancy, and from mothers and children (umbilical cord) at birth [62, 63].

The establishment of MoBa and initial data collection was based on a license from the Norwegian Data Protection Agency and approval from the Regional Committees for Medical and Health Research Ethics (REC).

MoBa is currently regulated by the Norwegian Health Registry Act. All MoBa participants have given their written informed consent to participate in the cohort study. The current study has been approved by REC South East Norway (REC reference: 23,136, 2014/163). All data are de-identified, and the linkage between MoBa and the different health registries was handled by NIPH along with the relevant registries.

Study design and measurements

The MoBa biobank contains 90,000 cord blood samples drawn at birth [63]. In our original study using the 450 k platform, we selected 384 samples from the biobank, and in the study using the EPIC platform, we selected 261 samples. Out of these samples, 611 samples were unique to either the 450 k data set or the EPIC data set, and 17 samples were measured on both the 450 k and EPIC platforms. The samples were selected based on prenatal exposure to paracetamol and child ADHD diagnosis, and all samples were term births (≥ 37 weeks). The 17 samples available in both data sets were all prenatally long-term exposed to paracetamol and had received an ADHD diagnosis.

Long-term prenatal exposure to paracetamol (Anatomical Therapeutic Chemical [ATC] code: N02BE01) was defined as the use of paracetamol for ≥ 20 days during pregnancy (coded as “yes” or “no”), based on a threshold from previous studies [64–68]. Use was self-reported and collected from three MoBa questionnaires (Q1, Q3 and Q4). Offspring diagnosis of ADHD was retrieved from the NPR (2008), containing all individual diagnoses asserted by specialists according to the 10th revision of the International Classification of Disease (ICD-10), as reported by governmental hospitals and outpatient clinics. Children were defined as having ADHD if they had received an ICD-10 diagnosis of hyperkinetic disorder (HKD; F90.0, F90.1, F90.8 or F90.9) between 2008 and 2016. HKD corresponds to ADHD in the Diagnostic and Statistical Manual (DSM) system [69–72], as an HKD diagnosis requires both inattentiveness and hyperactivity symptoms.

DNA methylation

Generation of DNAm data

The 450 k DNAm data from the samples in our original study are described elsewhere [21]. The samples assessed on the Infinium HumanMethylation EPIC BeadChip (Illumina) were processed similar to the 450 k data set [21]. Samples were randomly allocated to sample plates and beadchips, as previously described [21].

Quality control and pre-processing

Analyses were performed in the R programming language (<http://www.r-project.org/>). Quality control, normalisation and filtering of the data (Table 1) were performed using the default pipeline of *ENmix* [34]. The EPIC and 450 k data sets were pre-processed separately and all samples were included in the pre-processing ($n_{\text{EPIC}} = 261$; $n_{450\text{k}} = 384$). Subsequently, the 17 samples with repeated, cross-platform measurements were used for further analyses.

First, samples with >5% low-quality CpGs or low bisulphite intensity were removed (7 samples from the 450 k data set and 0 samples from the EPIC data set). Then, CpGs with >5% low-quality values were also removed (5598 and 8947 CpGs from the 450 k and EPIC data sets, respectively). Background correction was performed using the *ENmix* exponential-truncated-normal out-of-band (oob) method [34], dye bias correction was executed using RELIC (REgression on Logarithm of Internal Control probes) [73], and probe-type correction was achieved using RCP (Regression of Correlated Probes) [74]. We removed probes with SNPs overlapping with the CpG interrogation site or the nucleotide extension site ($n_{\text{EPIC}} = 29,176$; $n_{450\text{k}} = 16,803$), cross-reactive probes ($n_{\text{EPIC}} = 14,921$; $n_{450\text{k}} = 21,563$) [36, 75–77] and probes on the sex chromosomes ($n_{\text{EPIC}} = 17,532$; $n_{450\text{k}} = 10,012$). These pre-processing steps resulted in a total of 795,515 probes in the EPIC data set and 431,536 probes in the 450 k data set. Of these, 397,813 CpGs overlapped between the two platforms.

Pre-processing using the default settings of common analysis pipelines

The raw data were also pre-processed using the default settings of four other common EWAS analysis pipelines: *ChAMP* [50, 51], *minfi* [52], *RnBeads* [53] and *wateRmelon* [54]. Additionally, we used the default *RnBeads* pipeline [53], but changed the background and probe-type correction methods to *Enmix.oob* [34] and BMIQ [55], respectively. The CpGs were annotated based on *ilm10b4.hg19* [78].

Table 1 Overview of retained probes upon filtering of data from the EPIC and 450 k microarray platforms

	EPIC probes	450 k probes
Raw data	866,091	485,512
> 5% low-quality values	857,144	479,914
SNP-enriched probe removal	827,968	463,111
Cross-reactive probe removal	813,047	441,548
Sex chromosome removal	795,515	431,536

Statistical analyses

The β values (the ratio of methylated signal to the sum of methylated and unmethylated signal) were used for visualisations and calculation of all concordance measurements. To test for differences in distributions, we used the Kolmogorov–Smirnov test, and to test for differences in proportions, we used the Pearson's Chi-squared test. To examine the correlations between both samples and CpGs from the different microarrays, we estimated the Pearson's correlation coefficient. The ICC of each CpG was computed using the *irr* package [79]. We estimated the ICC by fitting an absolute agreement and mean of k raters ($k = 2$), two-way random effects model, as has previously been suggested for such comparisons [31]. The visualisation of the overlaps between studies of CpGs with mean DNAm differences >0.5 across platforms was generated using the *UpSetR* package [80]

Abbreviations

ADHD: Attention deficit/hyperactivity disorder; ATC: Anatomical therapeutic chemical; DSM: Diagnostic and statistical manual; CpG: 5'-Cytosine-phosphate-guanine-3' site; DNAm: DNA methylation; EWAS: Epigenome-wide association study; FDR: False discovery rate; HKD: Hyperkinetic disorders; ICC: Intra-class correlation coefficient; ICD-10: The 10th revision of the International Classification of Disease; MBRN: The Medical Birth Registry of Norway; MoBa: The Norwegian Mother, Father and Child Cohort Study; NIPH: The Norwegian Institute for Public Health; NPR: The Norwegian Patient Registry; PCA: Principal component analysis; PPV: Positive predictive value; REC: The Regional Committees for Medical and Health Research Ethics.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13148-022-01299-3>.

Additional file 1: Fig. S1. Heatmap and clustering of genotyping probes. Fig. S2. Overlap of the CpGs exhibiting differences in mean DNAm > 0.5 in five studies. Fig. S3. Probe types and density distribution of Pearson's and intra-class correlation coefficients. Fig. S4. Annotation groups and relation to Pearson's and intra-class correlation coefficient categories. Fig. S5. Histogram of the distribution of the per-CpG island intra-class correlation coefficients. Fig. S6. Density plot of DNAm levels stratified by annotation categories. Fig. S7. Histograms of the distribution of intra-class correlation coefficients for randomly paired samples. Fig. S8. Scatter plot of the difference in mean DNAm level against the intra-class correlation coefficient. Fig. S9. ICC distributions for the significant CpGs of our original study. Table S1. Overview of common pipelines with default settings for analysing DNA methylation data.

Additional file 2: Table of CpGs with corresponding intra-class correlation coefficients, Pearson's correlation coefficients, mean difference across platforms, and whether the CpG exhibited an ICC ≥ 0.5 both in the current study and in the Sugden et al. study.

Acknowledgements

The Norwegian Mother, Father and Child Cohort Study is supported by the Norwegian Ministry of Health and Care Services and the Ministry of Education and Research. We are grateful to all the participating families in Norway taking part in this ongoing cohort study.

Author contributions

EWO, HMEN and KG conceived the idea of and designed the study. EWO conducted the analyses. EWO generated plots and tables and drafted the first

version of the paper. EWO, HMEN, GKS, RL and KG all revised the paper. All authors read and approved the final manuscript.

Funding

This work was supported by the European Research Council (ERC) Starting Grant “DrugsInPregnancy” (grant number: 678033) and the Research Council of Norway (NFR) “Medications during pregnancy: effects on neurodevelopmental disorders in childhood and epigenetic outcomes” (grant number: 241117). MoBa is supported by the Norwegian Ministry of Health and the Ministry of Education and Research, NIH/NIEHS (contract number: NO1-ES-75558), NIH/NINDS (grant numbers: UO1 NS 047537-01 and UO1 NS047537-06A1) and the Norwegian Research Council/FUGE (grant numbers: 151918/S10).

Availability of data and materials

The data that support the findings of this study are available from the Norwegian Mother, Father and Child Cohort Study, but restrictions apply to the availability of these data and so are not publicly available. However, data are available from the authors upon reasonable request and with permission from the Norwegian Mother, Father and Child Cohort Study.

Declarations

Ethics approval and consent to participate

The establishment of MoBa and initial data collection was based on a license from the Norwegian Data Protection Agency and approval from the Regional Committees for Medical and Health Research Ethics (REC). MoBa is currently regulated by the Norwegian Health Registry Act. All MoBa participants have given their written informed consent to participate in the cohort study. The current study has been approved by REC South East Norway (REC reference: 23136, 2014/163). All data are de-identified, and the linkage between MoBa and the different health registries was handled by NIPH along with the relevant registries.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Pharmacoepidemiology and Drug Safety Research Group, Department of Pharmacy, Faculty of Mathematics and Natural Sciences, University of Oslo, Oslo, Norway. ²PharmaTox Strategic Research Initiative, Faculty of Mathematics and Natural Sciences, University of Oslo, Oslo, Norway. ³Department of Child Health and Development, Norwegian Institute of Public Health, Oslo, Norway. ⁴Department of Informatics, Faculty of Mathematics and Natural Sciences, University of Oslo, Oslo, Norway. ⁵Department of Medical Genetics, Oslo University Hospital and University of Oslo, Oslo, Norway. ⁶Centre for Fertility and Health, Norwegian Institute of Public Health, Oslo, Norway. ⁷Department of Research and Innovation, Division of Clinical Neuroscience, Oslo University Hospital, Oslo, Norway.

Received: 5 March 2022 Accepted: 12 June 2022

Published online: 28 June 2022

References

1. Mill J, Heijmans BT. From promises to practical strategies in epigenetic epidemiology. *Nat Rev Genet*. 2013;14:585–94.
2. Relton CL, Smith DG. Is epidemiology ready for epigenetics? *Int J Epidemiol*. 2012;41:5–9.
3. Godfrey KM, Lillycrop KA, Burdige GC, Gluckman PD, Hanson MA. Epigenetic mechanisms and the mismatch concept of the developmental origins of health and disease. *Pediatr Res*. 2007;61:5–10.
4. Barouki R, Mel n E, Herceg Z, Beckers J, Chen J, Karagas M, et al. Epigenetics as a mechanism linking developmental exposures to long-term toxicity. *Environ Int*. 2018;114:77–86.
5. Hjorth S, Bromley R, Ystrom E, Lupattelli A, Spigset O, Nordeng H. Use and validity of child neurodevelopment outcome measures in studies on prenatal exposure to psychotropic and analgesic medications—a systematic review. *PLoS ONE*. 2019;14(7):e0219778. <https://doi.org/10.1371/journal.pone.0219778>.
6. Kellogg M, Meador KJ. Neurodevelopmental effects of antiepileptic drugs. *Neurochem Res*. 2017;42:2065–70.
7. Bromley R, Weston J, Adab N, Greenhalgh J, Sanniti A, McKay AJ, et al. Treatment for epilepsy in pregnancy: neurodevelopmental outcomes in the child. *Cochrane Database Syst Rev*. 2014. <https://doi.org/10.1002/14651858.CD010236.pub2>.
8. Bauer AZ, Kriebel D, Herbert MR, Bornehag CG, Swan SH. Prenatal paracetamol exposure and child neurodevelopment: A review. *Horm Behav Academic Press*. 2018;101:125–47.
9. El Marroun H, White T, Verhulst FC, Tiemeier H. Maternal use of antidepressant or anxiolytic medication during pregnancy and childhood neurodevelopmental outcomes: a systematic review. *Eur Child Adolesc Psychiatry*. 2014;23:973–92.
10. Prady SL, Hanlon I, Fraser LK, Mikocka-Walus A. A systematic review of maternal antidepressant use in pregnancy and short- and long-term offspring’s outcomes. *Arch Womens Ment Health*. 2018;21:127.
11. Tosato S, Albert U, Tomassi S, Iasevoli F, Carmassi C, Ferrari S, et al. A Systematized review of atypical antipsychotics in pregnant women: balancing between risks of untreated illness and risks of drug-related adverse effects. *J Clin Psychiatry J Clin Psychiatry*. 2017;78:e477–89.
12. Ornoy A, Weinstein-Fudim L, Ergaz Z. Antidepressants, antipsychotics, and mood stabilizers in pregnancy: what do we know and how should we treat pregnant women with depression. *Birth Defects Res Birth Defects Res*. 2017;109:933–56.
13. Yazdy MM, Desai RJ, Brogly SB. Prenatal exposures and short and long term developmental outcomes: prescription opioids in pregnancy and birth outcomes: a review of the literature. *J Pediatr Genet*. 2015;4:56.
14. Radhakrishna U, Vishweswaraiah S, Uppala LV, Szymanska M, Macknis J, Kumar S, et al. Placental DNA methylation profiles in opioid-exposed pregnancies and associations with the neonatal opioid withdrawal syndrome. *Genomics*. 2021;113:1127–35.
15. Kallak TK, Br nn E, Fransson E, Johansson  , Lager S, Comasco E, et al. DNA methylation in cord blood in association with prenatal depressive symptoms. *Clin Epigenetics BioMed Central Ltd*. 2021;13:1–14.
16. Wachman EM, Wang A, Isley BC, Boateng J, Beierle JA, Hansbury A, et al. Placental OPRM1 DNA methylation and associations with neonatal opioid withdrawal syndrome, a pilot study. *Explor Med*. 2020;1:124–35.
17. Czamara D, Dieckmann L, R h S, Kraemer S, Rancourt RC, Sammallahti S, et al. Betamethasone administration during pregnancy is associated with placental epigenetic changes with implications for inflammation. *Clin Epigenetics BioMed Central Ltd*. 2021;13:1–17.
18. Vidal AC, Murphy SK, Murtha AP, Schildkraut JM, Soubry A, Huang Z, et al. Associations between antibiotic exposure during pregnancy, birth weight and aberrant methylation at imprinted genes among offspring. *Int J Obes (Lond)*. 2013;37:907.
19. Oni-Orisan OO, Dansereau LM, Marsit CJ, Smith LM, Neal CR, Della Grotta SA, et al. DNA methylation in children with prenatal methamphetamine exposure and environmental adversity. *Pediatr Res Pediatr Res*. 2021;89:1152–6.
20. Olstad EW, Nordeng HME, Gervin K. Prenatal medication exposure and epigenetic outcomes: a systematic literature review and recommendations for prenatal pharmacoeconomic studies. *Epigenetics*. 2021;
21. Gervin K, Nordeng H, Ystrom E, Reichborn-Kjennerud T, Lyle R. Long-term prenatal exposure to paracetamol is associated with DNA methylation differences in children diagnosed with ADHD. *Clin Epigenetics*. 2017. <https://doi.org/10.1186/s13148-017-0376-9>.
22. Heijmans BT, Mill J. Commentary: the seven plagues of epigenetic epidemiology. *Int J Epidemiol*. 2012;41:74–8.
23. Jones MJ, Moore SR, Kober MS. Principles and challenges of applying epigenetic epidemiology to psychology. *Annu Rev Psychol Annu Rev Psychol*. 2018;69:459–85.
24. Ryan J, Mansell T, Fransquet P, Saffery R. Does maternal mental well-being in pregnancy impact the early human epigenome? *Epigenomics*. 2017;9:313–32.
25. Campagna MP, Xavier A, Lechner-Scott J, Maltby V, Scott RJ, Butzkueven H, Jokubaitis VG, Lea RA. Epigenome-wide association studies: current

- knowledge, strategies and recommendations. *Clin Epigenetics*. 2021. <https://doi.org/10.1186/s13148-021-01200-8>.
26. Bose M, Wu C, Pankow JS, Demerath EW, Bressler J, Fornage M, et al. Evaluation of microarray-based DNA methylation measurement using technical replicates: the atherosclerosis risk in communities (ARIC) study. *BMC Bioinform BioMed Central Ltd*. 2014;15:1–10.
 27. Shvetsov YB, Song MA, Cai Q, Tiirikainen M, Xiang YB, Shu XO, et al. Intra-individual variation and short-term temporal trend in DNA methylation of human blood. *Cancer Epidemiol Biomarkers Prev*. 2015;24:490.
 28. Dugué PA, English DR, MacInnis RJ, Jung CH, Bassett JK, Fitzgerald LM, et al. Reliability of DNA methylation measures from dried blood spots and mononuclear cells using the HumanMethylation450k BeadArray. *Sci Rep. Nature Publishing Group*; 2016;6.
 29. Forest M, O'Donnell KJ, Voisin G, Gaudreau H, MacIsaac JL, McEwen LM, et al. Agreement in DNA methylation levels from the Illumina 450K array across batches, tissues, and time. *Epigenetics*. 2018;13:19.
 30. Xu Z, Taylor JA. Reliability of DNA methylation measures using Illumina methylation BeadChip. *Epigenetics Epigenetics*. 2021;16:495–502.
 31. Sugden K, Hannon EJ, Arseneault L, Belsky DW, Corcoran DL, Fisher HL, et al. Patterns of Reliability: Assessing the Reproducibility and Integrity of DNA Methylation Measurement. *Patterns*. 2020;1:100014.
 32. Logue MW, Smith AK, Wolf EJ, Maniates H, Stone A, Schichman SA, et al. The correlation of methylation levels measured using Illumina 450K and EPIC BeadChips in blood samples. *Epigenomics Future Sci Group*. 2017;9:1363–71.
 33. Zaimi I, Pei D, Koestler DC, Marsit CJ, De Vivo I, Tworoger SS, et al. Variation in DNA methylation of human blood over a 1-year period using the Illumina MethylationEPIC array. *Epigenetics*. 2018;13:1056–71.
 34. Zongli X, Niu L, Li L, Taylor JA. ENmix: a novel background correction method for Illumina HumanMethylation450 Beadchip. *Nucleic Acids Res*. 2016;44(3):e20–e20. <https://doi.org/10.1093/nar/gkv907>.
 35. Xu Z, Niu L, Taylor JA. The ENmix DNA methylation analysis pipeline for Illumina beadchip and comparisons with seven other preprocessing pipelines *clin epigenetics* 2021 131. *BioMed Central*. 2021;13:1–8.
 36. Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P, Van Dijk S, Muhlhäuser B, Stirzaker C, Clark SJ. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol*. 2016. <https://doi.org/10.1186/s13059-016-1066-1>.
 37. Solomon O, MacIsaac J, Quach H, Tindula G, Kobor MS, Huen K, et al. Comparison of DNA methylation measured by Illumina 450K and EPIC BeadChips in blood of newborns and 14-year-old children. *Epigenetics*. 2018;13:655–64.
 38. Fernandez-Jimenez N, Allard C, Bouchard L, Perron P, Bustamante M, Bilbao JR, et al. Comparison of Illumina 450K and EPIC arrays in placental DNA methylation. *Epigenetics*. 2019;14:1177–82.
 39. Cheung K, Burgers MJ, Young DA, Cockell S, Reynard LN. Correlation of infinium humanmethylation450K and METHYLATIONEPIC beadchip arrays in cartilage. *Epigenetics*. 2020;15:594–603.
 40. Moran S, Arribas C, Esteller M. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. *Epigenomics Future Med Ltd*. 2016;8:389–99.
 41. Han L, Zhang H, Kaushal A, Rezwan FI, Kadalayil L, Karmaus W, et al. Changes in DNA methylation from pre- to post-adolescence are associated with pubertal exposures. *Clin Epigenetics BioMed Central Ltd*. 2019;11:1–14.
 42. Li S, Nguyen TL, Wong EM, Dugué PA, Dite GS, Armstrong NJ, et al. Genetic and environmental causes of variation in epigenetic aging across the lifespan. *Clin Epigenetics BioMed Central Ltd*. 2020;12:1–12.
 43. Ewing E, Kular L, Fernandes SJ, Karathanasis N, Lagani V, Ruhmann S, et al. Combining evidence from four immune cell types identifies DNA methylation patterns that implicate functionally distinct pathways during Multiple Sclerosis progression. *EBioMedicine*. 2019;43:411–23.
 44. Neumann A, Walton E, Alemany S, Cecil C, González JR, Jima DD, et al. Association between DNA methylation and ADHD symptoms from birth to school age: a prospective meta-analysis. *Transl Psychiatry* 101. 2020;10:1–11.
 45. Felix JF, Joubert BR, Baccarelli AA, Sharp GC, Almqvist C, Annesi-Maesano I, et al. Cohort profile: pregnancy and childhood epigenetics (PACE) consortium. *Int J Epidemiol Oxford Academic*. 2018;47:22–23u.
 46. Breton CV, Marsit CJ, Faustman E, Nadeau K, Goodrich JM, Dolinoy DC, et al. Small-magnitude effect sizes in epigenetic end points are important in children's environmental health studies: the children's environmental health and disease prevention research center's epigenetics working group. *Environ Health Perspect Environ Health Perspect*. 2017;125:511–26.
 47. Dugué PA, English DR, MacInnis RJ, Joo JE, Jung CH, Milne RL. The repeatability of DNA methylation measures may also affect the power of epigenome-wide association studies. *Int J Epidemiol Int J Epidemiol*. 2015;44:1460–1.
 48. Koo TK, Li MY. A Guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15:155.
 49. Chen J, Just AC, Schwartz J, Hou L, Jafari N, Sun Z, et al. CpGFilter: model-based CpG probe filtering with replicates for epigenome-wide association studies. *Bioinformatics*. 2016;32:469.
 50. Morris TJ, Butcher LM, Feber A, Teschendorff AE, Chakravarthy AR, Wojdacz TK, et al. ChAMP: 450k Chip analysis methylation pipeline. *Bioinform Oxford Academic*. 2014;30:428–30.
 51. Tian Y, Morris TJ, Webster AP, Yang Z, Beck S, Feber A, et al. ChAMP: updated methylation analysis pipeline for Illumina BeadChips. *Bioinform Oxford Academic*. 2017;33:3982–4.
 52. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*. 2014;30(10):1363–9. <https://doi.org/10.1093/bioinformatics/btu049>.
 53. Müller F, Scherer M, Assenov Y, Lutsik P, Walter J, Lengauer T, et al. RnBeads 2.0: Comprehensive analysis of DNA methylation data. *Genome Biol. BioMed Central Ltd*. 2019;20:55.
 54. Pidsley R, Y Wong CC, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data *BMC Genomics. BioMed Central*. 2013;14:1–10.
 55. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinform Bioinform*. 2013;29:189–96.
 56. Dugué PA, Jung CH, Joo JE, Wang X, Wong EM, Makalic E, et al. Smoking and blood DNA methylation: an epigenome-wide association study and assessment of reversibility. *Epigenetics*. 2020;15:358.
 57. Bujang M, Baharum N. A simplified guide to determination of sample size requirements for estimating the value of intraclass correlation coefficient: a review. *Arch Orofac Sci*. 2017.
 58. Li M, Zou D, Li Z, Gao R, Sang J, Zhang Y, Li R, Xia L, Zhang T, Niu G, Bao Y, Zhang Z. EWAS Atlas: a curated knowledgebase of epigenome-wide association studies. *Nucleic Acids Res*. 2019;47(D1):D983–8. <https://doi.org/10.1093/nar/gky1027>.
 59. Battram T, Yousefi P, Crawford G, Prince C, Babaei MS, Sharp G, et al. The EWAS Catalog: a database of epigenome-wide association studies. *OSF Prepr. OSF Preprints*; 2021;4.
 60. Magnus P, Irgens LM, Haug K, Nystad W, Skjærven R, Stoltenberg C, et al. Cohort profile: The Norwegian Mother and Child Cohort Study (MoBa). *Int J Epidemiol Oxford Academic*. 2006;35:1146–50.
 61. Magnus P, Birke C, Vejrup K, Haugan A, Alsaker E, Daltveit AK, et al. Cohort profile update: The Norwegian mother and child cohort study (MoBa). *Int J Epidemiol Int J Epidemiol*. 2016;45:382–8.
 62. Paltiel L, Haugan A, Skjærden T, Harbak K, Bækken S, Stensrud NK, et al. The biobank of the Norwegian mother and child cohort study – present status. *Nor Epidemiol Norwegian Epidemiol Soc*. 2014;24:29–35.
 63. Rønningen KS, Paltiel L, Meltzer HM, Nordhagen R, Lie KK, Hovengen R, et al. The biobank of the Norwegian mother and child cohort study: a resource for the next 100 years. *Eur J Epidemiol Eur J Epidemiol*. 2006;21:619–25.
 64. Brandlistuen RE, Ystrom E, Nulman I, Koren G, Nordeng H. Prenatal paracetamol exposure and child neurodevelopment: A sibling-controlled cohort study. *Int J Epidemiol*. 2013;42:1702–13.
 65. Stergiakouli E, Thapar A, Smith GD. Association of acetaminophen use during pregnancy with behavioral problems in childhood: evidence against confounding. *JAMA Pediatr Am Med Assoc*. 2016;170:964–70.
 66. Avella-García CB, Julvez J, Fortuny J, Rebordosa C, García-Esteban R, Riano Galán I, et al. Acetaminophen use in pregnancy and neurodevelopment: attention function and autism spectrum symptoms. *Int J Epidemiol*. 2016;45:1987–96.

67. Thompson JMD, Waldie KE, Wall CR, Murphy R, Mitchell EA. Associations between acetaminophen use during pregnancy and ADHD symptoms measured at ages 7 and 11 years. *PLoS ONE*. 2014;9(9):e108210. <https://doi.org/10.1371/journal.pone.0108210>.
68. Liew Z, Ritz B, Rebordosa C, Lee PC, Olsen J. Acetaminophen use during pregnancy, behavioral problems, and hyperkinetic disorders. *JAMA Pediatr Am Med Assoc*. 2014;168:313–20.
69. World Health Organization. Chapter V: Mental and Behavioural Disorders. 10th Revis Int Stat Classif Dis Relat Heal Probl. 2018.
70. American Psychiatric Association. Neurodevelopmental Disorders. Diagnostic Stat Man Ment Disord. American Psychiatric Association; 2013.
71. Thapar A, Pine DS, Leckman JF, Scott S, Snowling MJ, Taylor E. Rutter's Child and Adolescent Psychiatry: Sixth Edition. Rutter's Child Adolesc. Psychiatry Sixth Ed. Chichester, UK: John Wiley & Sons, Ltd; 2015.
72. Wieselberg HM, Rutter M, Thorley G. Which boys respond to stimulant medication? a controlled trial of methylphenidate in boys with disruptive behaviour. *Psychol Med Psychol Med*. 1987;17:121–43.
73. Zongli Xu, Langie SAS, De Boever P, Taylor JA, Niu L. RELIC: a novel dye-bias correction method for Illumina Methylation BeadChip. *BMC Genomics*. 2017. <https://doi.org/10.1186/s12864-016-3426-3>.
74. Niu L, Xu Z, Taylor JA. RCP: a novel probe design bias correction method for Illumina Methylation BeadChip. *Bioinformatics Bioinformatics*. 2016;32:2659–63.
75. Benton MC, Johnstone A, Eccles D, Harmon B, Hayes MT, Lea RA, et al. An analysis of DNA methylation in human adipose tissue reveals differential modification of obesity genes before and after gastric bypass and weight loss. *Genome Biol BioMed Central*. 2015;16:1–21.
76. Chen Y, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics* 8. 2013;8:203–9. <https://doi.org/10.4161/epi.23470>.
77. McCartney DL, Walker RM, Morris SW, McIntosh AM, Porteous DJ, Evans KL. Identification of polymorphic and off-target probe binding sites on the Illumina Infinium MethylationEPIC BeadChip. *Genomics Data. Elsevier Inc*; 2016;9:22–4.
78. Hansen KD. IlluminaHumanMethylationEPICanno.ilm10b4.hg19: Annotation for Illumina's EPIC methylation arrays [Internet]. 2017.
79. Gamer M, Lemon J, Fellows I, Singh P. Irr: Various coefficients of interrater reliability and agreement. R package version, 0.84.1. 2019.
80. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics Oxford Academic*. 2017;33:2938–40.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

