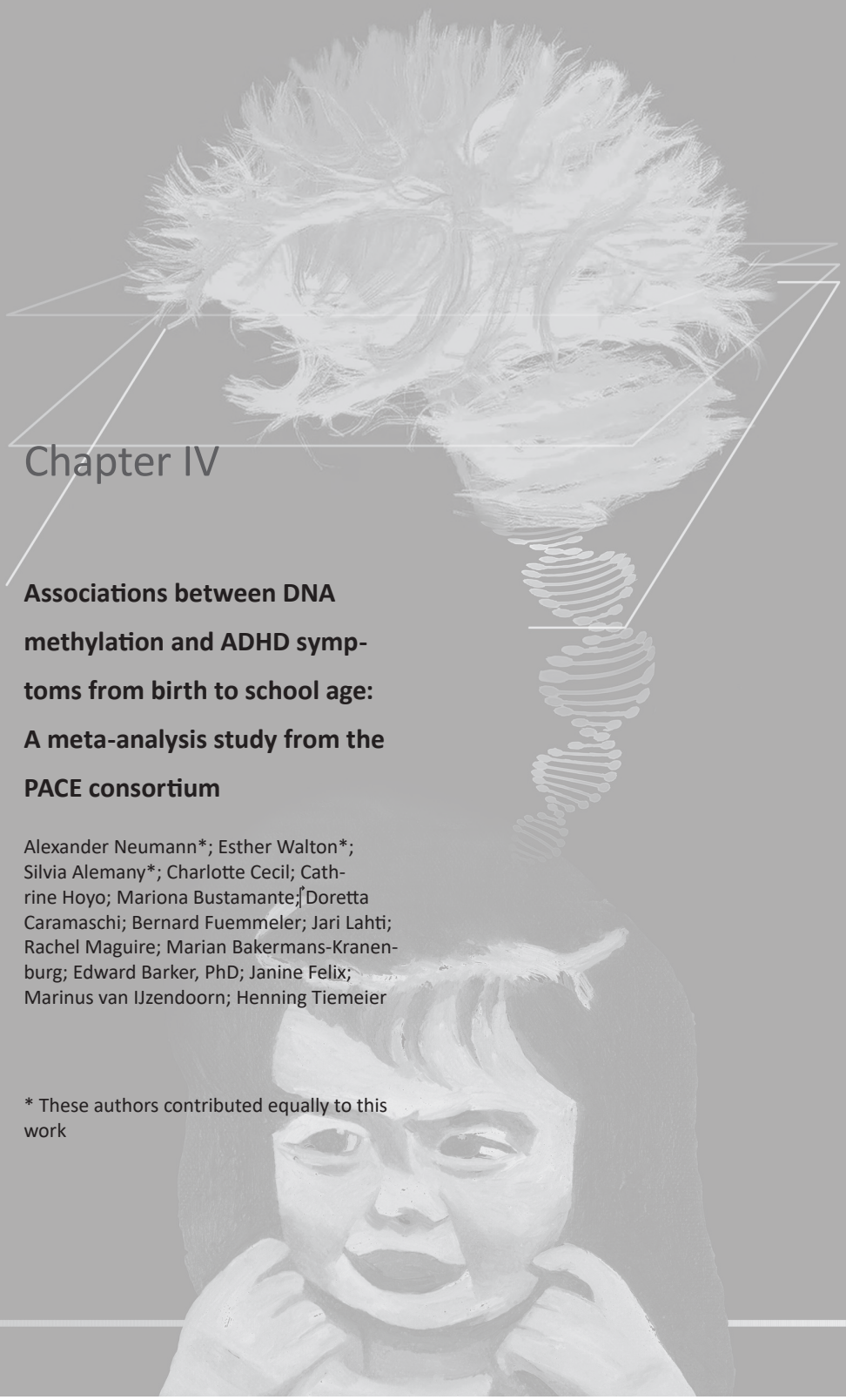

Chapter IV

Associations between DNA methylation and ADHD symptoms from birth to school age: A meta-analysis study from the PACE consortium

Alexander Neumann*; Esther Walton*; Silvia Alemany*; Charlotte Cecil; Cathrine Hoyo; Mariona Bustamante; Doretta Caramaschi; Bernard Fuemmeler; Jari Lahti; Rachel Maguire; Marian Bakermans-Kranenburg; Edward Barker, PhD; Janine Felix; Marinus van IJzendoorn; Henning Tiemeier

* These authors contributed equally to this work

ABSTRACT

Attention-deficit and hyperactivity disorder (ADHD) is a common childhood disorder with a substantial genetic component. However, to what extent epigenetic mechanisms play a role in the etiology of the disorder is much less well-known. Previous studies have identified several DNA methylation sites associated with ADHD. Yet, large epigenome-wide analyses (EWAS) featuring multiple independent cohorts are lacking. We performed an EWAS within the Pregnancy And Childhood Epigenetics (PACE) Consortium to identify DNA methylation sites associated with ADHD symptoms, the first prospective meta-analytic EWAS in child psychiatry. As DNA methylation changes over time, we performed two EWAS at two methylation assessment periods: birth and school-age. We examined associations of DNA methylation in cord blood with repeatedly assessed ADHD symptoms (age range 4-15 years) in 2477 children from five cohorts and DNA methylation at school-age (age 7-9 years) with concurrent ADHD symptoms (age 7-11 years) in 2374 children from ten cohorts. The regression estimates correlated with 0.30 between both time points, after exclusion of dependent samples, suggesting that the association between DNA methylation and ADHD is to some extent age independent. At birth, we identified 9 probes that were associated with later ADHD symptoms. Peripheral DNA methylation in only one of these probes correlated consistently with brain methylation. This probe (cg01271805) lies in the promotor region of ERC2, which regulates neurotransmitter release. Another genome-wide significant probe (cg25520701) lies within the gene CREB5, which is associated with neurite outgrowth and its genetic variants were previously related to ADHD. In contrast, no probes reached genome-wide significance when ADHD was associated with school-age DNA methylation, indicating that the methylation profiles of ADHD have higher explanatory power at birth. In conclusion, the results suggest that DNA methylation at birth may hold promise as a prognostic indicator for ADHD risk, but future studies are needed to confirm the utility as biomarker and presence of causal pathways.

INTRODUCTION

Attention-deficit and hyperactivity disorder (ADHD) is a common disorder characterized by age-inappropriate impulsivity, excessive activity and attention problems. Symptoms often become apparent during school-age with a world-wide prevalence of 5- 7.5%.[1] Genetic heritability is estimated between 64% and 88% in twin studies.[2,3] Additionally, several environmental factors are suspected to impact ADHD, such as prenatal maternal smoking or lead exposure.[4–7] However, it remains unclear through which pathways exactly genetics, and also the environment, affect ADHD risk. A possibility is that DNA methylation – an epigenetic mechanism that regulates gene expression – may mediate the effect of genetic and/or environmental determinants. DNA methylation can be affected by both environmental and genetic factors, it plays an essential role in healthy development, and in turn, disruptions in DNA methylation have been associated with disease risk, including psychiatric disorders. Several studies investigated DNA methylation in relation to ADHD diagnoses or symptoms using either a candidate gene approach or epigenome-wide assocation study (EWAS) in peripheral blood and saliva tissue, as reviewed previously[8,9]. A prominent hypothesis states that deficiencies in the dopamine system of the brain have an impact on ADHD development. This hypothesis is supported by imagining research, as well as observations that dopamine related genes are associated with ADHD risk.[4,10] Consequently, candidate gene studies investigating DNA methylation have primarily focused on genes related to dopamine function. For instance, DNA methylation in SLC6A3[11], SLC6A4[12], DRD4[12–14], DRD5[13], DAT1[13] genes have all been associated with ADHD diagnoses or ADHD symptoms, though not consistently[15]. Besides the candidate gene approach, two studies tested DNA methylation across the whole genome. One study performed an EWAS in school-aged children using a case-control design.[16] The study identified differentially methylated probes in the VIPR2 gene, a gene expressed in the caudate and previously associated with psychopathology. Another EWAS investigated DNA methylation repeatedly at birth and age 7.[17] This EWAS found 13 probes to be associated with ADHD trajectories from age 7 to 15, located in SKI, ZNF544, ST3GAL3 and PEX2. Interestingly, the methylation status of these probes at age 7 was not associated with ADHD. So far, there were no attempts to replicate the findings from these genome-wide studies.

While considerable research has begun to investigate DNA methylation in relation to ADHD occurrence, large multi-center epigenome-wide studies are lacking. A meta-analytic EWAS pooling results from independent research centers has several advantages. Large sample size increases power to detect probes, which otherwise would have been masked by the multiple testing correction for hundreds of thousands tests. The multi-center design also increases the generalizability of the results, as identified probes are more likely to show association under different research settings. Here, we performed the first epigenome-wide meta-analysis to identify DNA metyhlation sites

associated with ADHD symptoms. This meta-analysis was performed within the Pregnancy And Childhood Epigenetics (PACE) Consortium[18]. In contrast to the genome, the epigenome changes over time (due to environmental, genetic or stochastic effects) and is dependent on tissue type. To address sensitivity to timing, we tested DNA methylation at birth using cord blood and during school-age (age 7-9 years) using whole blood. In the analyses of methylation at birth, the aim was to predict the occurrence of ADHD symptoms at age 4-15. We took advantage of the fact that many participating cohorts assessed ADHD repeatedly and employed a repeated measures design to increase precision. Furthermore, we utilized data in childhood to examine cross-sectional DNA methylation patterns associated with ADHD symptoms at school age. With this design we are able to test for methylation effects, which are representative of prenatal exposures, (prospective birth analysis) as well as associations at later age, which reflect exposures until school-age or potentially consequences of ADHD symptoms. As for the second peculiarity of DNA methylation, the tissue dependence, we were only able to study peripheral tissue. However, we looked up in databases the correlation between blood and brain methylation to get a sense of whether any findings are exclusive to peripheral tissue.

METHODS

This study consists of two parts: the birth methylation EWAS and the school-age methylation EWAS which will be described successively.

Birth Cord Blood Methylation

Participants

Five cohorts (ALSPAC, GENR, INMA, NEST and PREDO) in the PACE consortium had information on DNA methylation in cord blood, and measured ADHD symptoms at later ages. These cohorts have a combined sample size of 2477. Participants had mostly European ancestry, with the exception of NEST, which also included participants with African ancestry. Participants with African and European ancestry were analyzed separately and treated as two separate studies in the meta-analysis.

DNA Methylation and QC

DNA Methylation was measured at birth using cord blood. The Illumina Infinium HumanMethylation450K BeadChip was used to interrogate CpG probes in all cohorts. Outlying methylation levels exceeding three times the interquartile range were removed before analysis. Each cohort ran an EWAS separately and results were then meta-analyzed centrally. The distribution of the regression estimates and p-values were examined for each study individually and for pooled results. Deviations from a normal

Table 1: Cohort characteristics of birth methylation EWAS

Cohort	Ancestry	n	ADHD Age	Instrument	33%	50%	66%	λ	Inflation	Bias
ALSPAC	European	714	7, 10, 12, 15	DAWBA	-0.21	0.25	0.89	1.60	1.10	0.37
GENR	European	1191	6,8,10	CBCL (6,10), Conners (8)	-0.48	0.01	0.53	1.51	1.20	0.05
INMA	European	325	7,9	Conners (7), CBCL (9)	-1.37	-0.40	0.43	0.80	0.87	-0.19
NEST	African	55	5	BASC	-3.50	-0.03	3.63	1.16	1.10	0.00
NEST	European	56	5	BASC	-2.54	-0.09	2.36	0.80	0.92	-0.01
PREDO	European	136	5	Conners	-1.55	-0.25	1.20	1.45	0.95	0.21
META	-	2477	-	-	-0.37	0.02	0.42	1.86	1.10	0.01

n Number of participants
33%, 50%, 66% Quartiles of regression coefficient distribution
λ Inflation of p-values
Inflation Inflation of p-values due to suspected bias
Bias Trend toward negative/positive distribution of regression coefficients due to suspected bias

distribution of regression estimates or a higher number of low p-values than expected under the null may be both signs of residual confounding, but may also be the result of a true signal. To help in interpretation of the results, we used the BACON method.[19] BACON analyzes the distribution of regression coefficients and estimates an empirical null distribution using a Bayesian approach. Results can then be compared against the empirical null, which already includes biases, rather than the theoretical null. After meta-analysis we excluded the CpG probes, that were available in less than four cohorts and fewer than 1000 participants, as well as allosomal probes, due to dosage compensation complicating interpretation of results.

ADHD Symptoms

ADHD symptoms were measured when children were 4-15 year old with parent-rated instruments, specifically the Behavior Assessment System for Children (BASC), Child Behavior Checklist (CBCL), Conners and The Development and Well-Being Assessment (DAWBA). If a cohort had ADHD symptoms measured repeatedly, every assessment wave was jointly analyzed in a mixed model (see statistical analysis). The repeated measure design increased power of the analysis by increasing precision of the ADHD severity estimate and by an increase of the sample size, since missing data in one or two of the assessments can be handled with maximum likelihood. Given the variety of instruments used within and across cohorts, all ADHD scores were z-score standardized to enable repeated measures analysis and meta-analysis.

Statistical analysis

Cohorts with repeated assessment ages were analyzed using linear mixed models. The outcome were z-scores of ADHD symptoms and the main predictor were methylation betas. Each CpG probe was analyzed separately and adjusted for multiple correction using Bonferroni adjustment. To account for dependence due to repeated measurements, we used a random intercept on the participant level. In addition, we used a random intercept on the batch level. The following potential confounders were included as fixed effects: maternal age, educational level, smoking status (yes vs no during pregnancy), child gestational age, gender, and cell proportions (Bakulski reference, a cord blood specific reference).[20] Mixed models were fitted using restricted maximum likelihood. Missing outcome data was handled with maximum likelihood, as long as a participant had at least one valid outcome. We used R[21] with the lme4[22] package to estimate the models. Cohorts with a single ADHD assessment wave used an equivalent linear regression model without random intercept on participant level. Batch effects in this case could be adjusted with fixed or random effects.

Meta-analysis was performed using the Han and Eskin random effects model.[23] This model does not assume that true effects are homogeneous between cohorts, however, it does assume that null effects are homogeneous. This modified version of the random

effect model has comparable power to a fixed effects analysis, while better accounting for study heterogeneity, such as ancestry differences, in simulation studies.[23]

Follow-up analyses

We performed several look-ups of genome-wide significant probes to better characterize findings. We used the BECon database26 to check the correlation between peripheral and brain methylation levels in post-mortem tissue. To characterize to which extent probes were under genetic influence we looked the hits up in MeQTL[24] and twin heritability databases.[25] We also attempted to replicate genome-wide significant probes reported in a previous EWAS performed in the ALSPAC cohort.[17] As this cohort also participated in this meta-analysis, we reran the meta-analysis with ALSPAC excluded to achieve an independent replication sample.

Pathway Analysis

We performed pathway enrichment analysis with the missMethylpackage[26] on probes showing suggestive evidence of association (P<1E-05). We used as references: gene ontology (GO), KEGG and curated gene sets (C2; http://software.broadinstitute.org/gsea/msigdb/collections.jsp#C2) from the Broad Institute Molecular signatures database[27]. missMethylpackage adjusts the p-values for the number of CpGs associated with each gene[26], since genes with larger numbers of probes are more likely to have significantly differentially methylated CpGs, biasing gene set analysis.[28] The packages correct by multiple testing using the false discovery rate method.

To test enrichment for regulatory features (gene relative position, CpG island relative position and blood chromatin states) we applied $\chi 2$ tests. Enrichment tests were performed for all CpGs, and for hypo and hypermethylated CpGs separately. CpG annotation was performed with the IlluminaHumanMethylation450kanno.ilmn-12.hg19 R package.[29] Annotation to 15 chromatin states was retrieved from 27 blood cell types from the Roadmap Epigenomics Project web portal (https://egg2.wustl.edu/roadmap/web_portal/). Each CpG in the array was annotated to one or several chromatin states by taking a state as present in those loci if it was described in at least 1 of the27 blood-related measurements.

School-age methylation

Participants

Four cohorts (ALSPAC, GENR, HELIX and GLAKU) with a combined sample size of 2374 joined the school-age methylation EWAS. Helix consists of six different sub-cohorts, which were pre-processed and analyzed jointly.[30] All cohorts had participants with European ancestry, except HELIX, which also included participants with a Pakistani

Table 2: Cohort characteristics of school-age methylation EWAS

Cohort	Ancestry	n	Methylation Age	ADHD Age	Instrument	33%	50%	66%	λ	Inflation	Bias
ALSPAC	European	651	7	7	DAWBA	-0.61	-0.10	0.54	1.09	1.00	**-0.08**
GENR	European	395	10	10	CBCL	-0.93	-0.00	0.98	1.00	0.97	**-0.01**
GLAKU	European	215	11	11	CBCL	-0.79	0.31	1.50	0.92	0.96	**0.13**
HELIX	European	1034	8	8	CBCL	-0.26	0.47	1.40	1.11	0.98	**0.28**
HELIX	Pakistani	79	8	8	CBCL	-1.66	1.86	5.48	0.98	0.96	**0.26**
Meta	-	2374	-	-	-	-0.24	0.14	0.62	0.96	0.92	**0.14**

n Number of participants
33%, 50%, 66% Quartiles of regression coefficient distribution
λ Inflation of p-values
Inflation Inflation of p-values due to suspected bias
Bias Trend toward negative/positive distribution of regression coefficients due to suspected bias

background. Participants of Pakistani national origin were treated as separate sample in the meta-analysis. See Table 2 for cohort characteristics.

DNA Methylation and QC

DNA methylation was measured at ages 7-11 using whole blood. The Illumina Infinium HumanMethylation450K BeadChip and Infinium MethylationEPIC Kit were used to interrogate CpG probes. QC steps were identical to the birth methylation EWAS.

ADHD Symptoms

ADHD symptoms were measured at the same age as DNA methylation at ages 7-11 years using the parent-rated measures DAWBA and CBCL.

Statistical analysis

The statistical model was similar to the linear regression model used in the birth methylation EWAS. However, cell counts were estimated with the Houseman reference31 as opposed to Bakulski, as Bakulski is specific to cord blood. We also added assessment age as covariate, because ADHD assesment age may correlate with DNA methylation assessment age, which in turn may be associated with methylation levels, which presents a confounding risk. The meta-analysis methods were identical to the birth methylation EWAS.

Follow-up analyses

As we did not find any genome-wide significant results (see Results) and observed an overall low signal, we did not perform follow-up anlayses. However, we did attempt to replicate six probes identified as most suggestive in a previous case-control EWAS, which assessed methylation and ADHD in school-age.[16]

RESULTS

Birth Cord Blood Methylation

EWAS Quality Check

We first examined whether the beta distribution of the individual cohorts was approximately normal with median regression estimates of 0, see Table 1. The distributions did not show signs of errors in analysis, however, one cohort (ALSPAC) showed a trend towards more positive estimates, whereas two others (INMA and PREDO) showed

more negative values. Furthermore, four out of the six cohorts showed a high λ, indicating larger number of low p-values than expected under the null.

To distinguish whether these trends are due to biases in the analyses, e.g. population stratification, or represent real methylome-wide effects, we estimated the inflation with the BACON method. This analysis suggested that the majority of the inflation is due to a true signal, as indicated by inflation values clearly lower than λ. Additional evidence that the inflation was not due to spurious associations was provided by a sensitivity analysis. Confounding due to population stratification or batch effects would likely be detectable at lower sample sizes, thus absence of inflation at lower sample sizes would indicate that inflation is due to power to detect associations not biased by these global strong effects. To test whether this is the case, we restricted the GenR sample randomly to 900 and 1100 participants, who had DNA methylation data available, which resulted in a maximum of 812 and 991 participants respectively due to missing covariates. In the case of 812 participants the lambda was 0.96, with 991 participants λ=1.21 and with the full sample of 1191 participants λ=1.51. Thus we concluded that the inflation is dependent on the power of the sample rather than spurious associations, which would also occur at lower sample size.

The BACON analyses also indicated a trend towards positive/negative values in some of the datasets, which can indicate confounding, e.g. by population stratification. However, all analyses were conducted in genetically homogoneous samples (or were stratified for ancestry), all cohorts adjusted for the same extensive list for possible confounders and all cohorts adjusted for batch effects. The addition of further variables, such as principal components of ancestry was tested in GENR and ALSPAC, but they did not meaningfully change results and were therefore not included.

We conducted the meta-analysis under the assumption that any such biases will be corrected in the pooled analysis, since they were not homogenous across cohorts. Indeed, the pooled estimates did not show a trend towards positive or negative regression estimates (Median=0.02), although showed overrepresentation of low p-values (λ=1.86, see QQ Plot in Figure 1). The BACON estimates for inflation, however, suggested that these are mostly due to a true signal (Inflation=1.1)

Single Probe Analysis

We performed a meta-analysis of cord blood EWAS results from six independent cohorts, pooling DNA methylation at birth in cord blood at 472,817 CpG sites. See Figure 2 for Manhattan plot. Nine CpG sites showed genome-wide significance at a Bonferroni correction threshold of $p < 1E-07$ ($5.24E-08 > p > 4.95E-09$). These probes predicted between 0.16SD and 0.415SD higher ADHD symptoms when the probe has 10% less methylation among all cells. The results therefore suggest that lack of methylation at these sites in the prenatal period is associated with higher number of ADHD symptoms in later life. Eight of the top probes were available in the BECon database[32]. According to the database, these eight probes are typically methylated in both whole

Table 3: EWAS Results

				Birth methylation					School-age methylation				
CpG	**Gene**	**Chr**	**Position**	**$n_{studies}$**	**n**	**B**	**SE**	**p**	**$n_{studies}$**	**n**	**B**	**SE**	**p**
cg25520701	CREB5	7	28800657	6	2450	-3.53	0.60	4.95E-09	5	2279	-0.13	1.09	0.94
cg24838839	Inter-genic	5	61031569	6	2468	-4.15	1.79	3.95E-08	5	2287	1.52	1.38	0.33
cg22997238	Inter-genic	7	36014218	6	2465	-1.63	0.30	8.81E-08	5	2291	-0.06	0.47	0.94
cg21600027	Inter-genic	4	124443502	6	2464	-3.04	0.81	2.64E-08	5	2281	0.98	0.89	0.33
cg17876201	ZBTB38	3	141139991	6	2457	-4.41	1.20	7.58E-09	4	2066	0.56	1.32	0.73
cg11251614	PPIL1	6	36839846	6	2451	-3.43	0.68	3.89E-08	5	2276	0.77	1.52	0.68
cg09762907	TRERF1	6	42290256	6	2460	-2.11	0.39	8.76E-08	5	2284	-0.55	0.64	0.46
cg09158638	Inter-genic	16	62309996	6	2470	-2.55	1.40	1.89E-08	5	2270	-0.33	1.04	0.80
cg01271805	ERC2	3	55694954	6	2469	-2.86	1.71	5.24E-08	5	2289	0.28	0.73	0.76

Chr Chromosome
$n_{studies}$ Number of studies
n Number of participants
B Regression coefficient
SE Standard error

Table 4: Replication of Walton et al. EWAS (ADHD trajectories and cord blood methylation)

				Discovery				Replication				
CpG	**Gene**	**Chr**	**Position**	**$n_{studies}$**	**n**	**B**	**p**	**$n_{studies}$**	**n**	**B**	**SE**	**p**
cg18587973	CDADC1	13	49822535	1	817	0.17	1.2E-06	5	1755	3.54	1.55	0.03
cg27469152	EPX	17	56282313	1	817	-0.18	2.0E-07	5	1763	-1.20	0.85	0.20
cg16290904	PEX2	8	77912348	1	817	0.17	7.4E-07	5	1761	-2.99	2.60	0.23
cg03905179	MAFK	7	1582588	1	817	0.17	1.3E-06	5	1756	0.70	1.67	0.46
cg24843380	ZNF454	5	178367827	1	817	0.17	1.6E-06	5	1762	-1.73	2.05	0.47
cg05653018	ELF3	1	201979533	1	817	0.17	1.4E-06	5	1763	0.45	0.60	0.53
cg15096815	JUN	1	59249838	1	817	-0.18	3.5E-07	5	1763	0.55	1.08	0.68
cg24481594	SKI	1	2190850	1	817	-0.20	1.5E-08	5	1763	-0.42	1.08	0.76
cg26263766	ZNF544	19	58739734	1	817	0.17	8.7E-07	5	1714	-0.31	1.13	0.84
cg01324543	CCDC30	1	42999439	1	817	-0.17	7.2E-07	5	1763	0.59	1.75	0.92
cg13714586	FBXW5	9	139838358	1	817	0.17	1.3E-06	5	1752	0.72	4.50	0.93
cg09989037	ST3GAL3	1	44300942	1	817	-0.17	9.5E-07	5	1763	-0.17	0.56	0.96
cg22193912	MAFG	17	79881523	1	817	0.17	1.3E-06	5	1763	0.24	0.81	0.99

Chr Chromosome
$n_{studies}$ Number of studies
n Number of participants
B Regression coefficient
SE Standard error

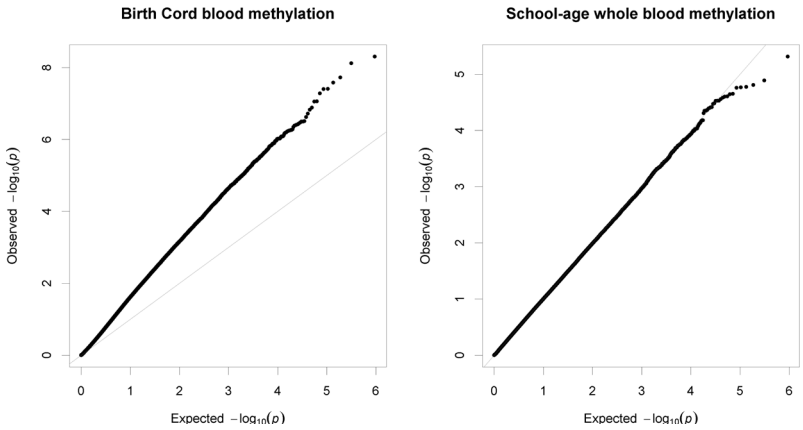

Figure 1: Quantile-quantile plot of observed -log 10 p values vs expected -log 10 p values assuming chance findings in birth methylation EWAS (left) and school-age methylation EWAS (right). Diagonal line indicates a p value distribution compatible with chance finding. Upward deviations indicate p values more significant than expected.

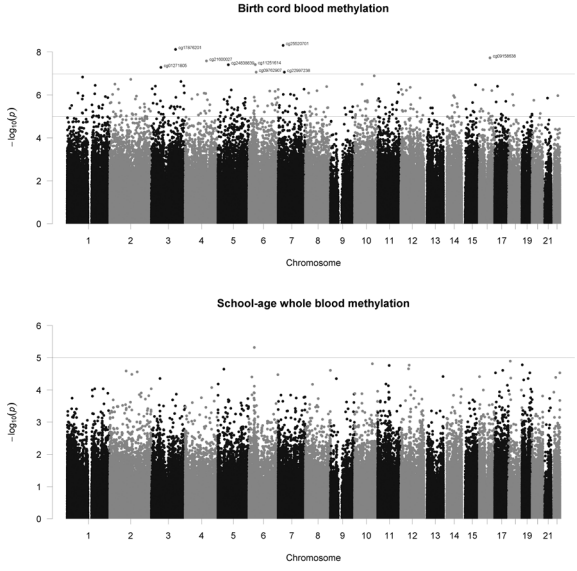

Figure 2: Manhattan plot of -log 10 p values vs CpG position for birth methylation EWAS (top) and school-age methylation EWAS (bottom).

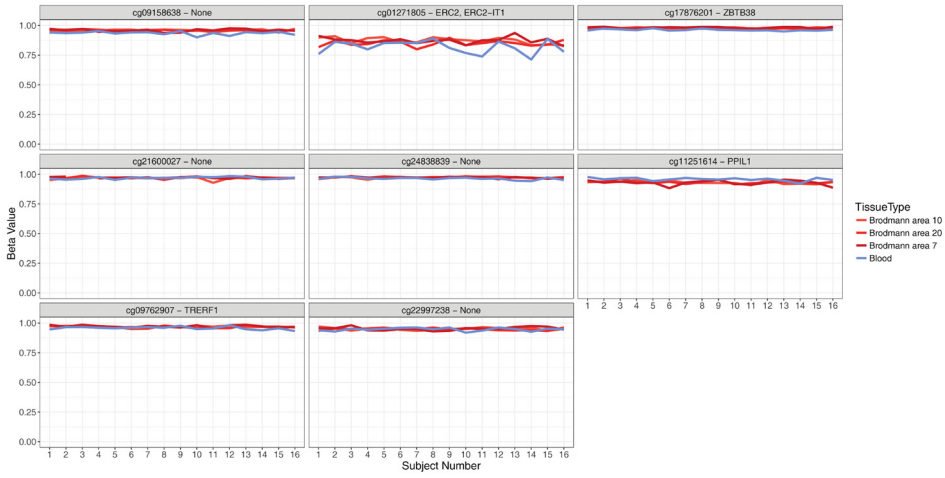

Figure 3: Methylation levels in blood and brain tissue in BECon database.

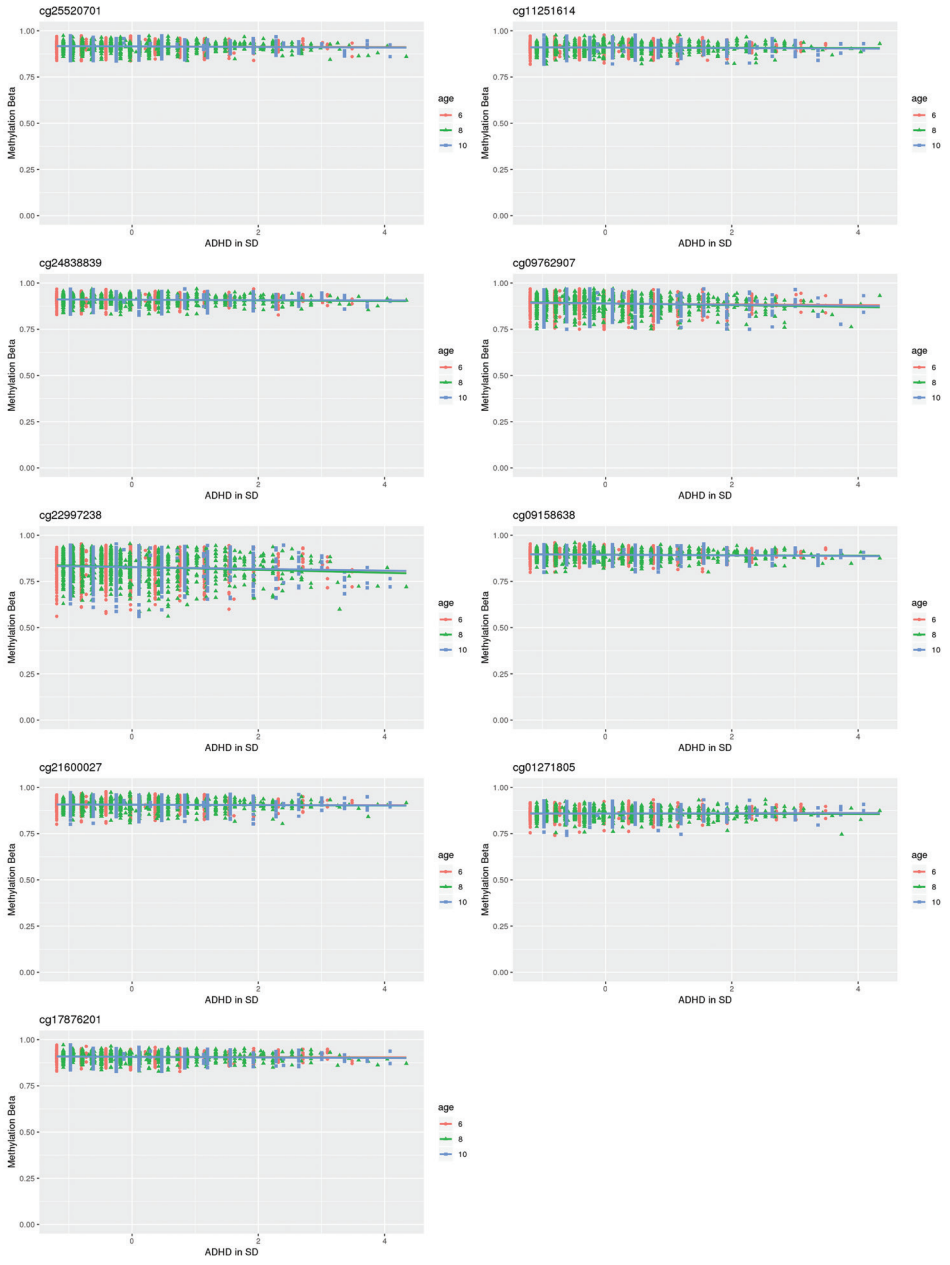

Figure 4: Scatter plot of methylation levels at birth dependent on levels of ADHD symptoms in the GenR cohort. Only methylation levels for genome-wide significant probes in the birth EWAS are shown.

Figure 5: Blood-brain correlation of methylation levels for genome-wide significant probes in the birth EWAS according to the BECon database.

blood as well as in the brain (see figure 3 and figure 5). A lookup in the BECon database revealed that the CpG site cg01271805 in the promoter region of gene ERC2 shows variable methylation in three brain regions (BA10, BA20, BA7). This may result in meaningful alteration in gene expression, as ERC2's expression profile is specific to brain tissues, with the highest expression in frontal cortex (BA9)[33]. Importantly, methylation levels in the brain are moderately correlated with whole blood methylation (0.33-0.46) (Figure 4), making peripheral cg01271805 levels a useful marker for brain methylation levels. The other hits showed less consistent correlations between blood and brain tissues and associated genes had less specificity for expression in the brain, based on GTEx data. No significant SNP was associated with our top hits when accounting for linkage disequilibrium with GCTA according to the MeQTL database[24]. Furthermore, no SNP had substantial twin heritability in a previous study.[25] After adjusting for inflation and bias with BACON, only one CpG remained statistically significant (cg25520701, CREB5, ß =-3.54, SE = 0.66, p = 9.59E-08). However, the p-values are based on a more conservative traditional random effects test rather than the modified Han and Eskin method.

Pathway Analysis

Two hundred forty nine probes showed suggestive (P<1E-05) association and were annotated to 182 unique genes. Among these probes no pathway survived multiple testing correction.

Suggestive CpGs, specially the hypomethylated, were enriched in intergenic regions. Suggestive hypomethylated probes were enriched for 3'UTR regions and depleted for TSS200 and first exon regions, open sea, north shelf and south shelf regions, south shore and islands. In regards to chromatin states, hypomethylated probes showed enrichment for transcription (Tx and TxWk) and quiescent positions and depletion for transcription start site positions (TSSA, TxFlnk, TxFlnk) and bivalent (EnhBiv) and repressor (ReprPC) positions. Hypermethylated probes showed the opposite chromatine state patterns.

Replication of previous EWAS

We attempted to replicate 13 CpG, which DNA methylation at birth was previously associated with ADHD trajectories.[17] However, no probe survived multiple-testing correction and effect directions were inconsistent between the two studies. (Table 4)

School-age methylation

EWAS Quality Checks

The beta regression distribution showed no signs of errors, but three of the cohorts showed a trend towards positive associations in separate analyses (Table 2). The lambda was below 1.11 for all cohorts. Further analysis of the cohorts with BACON suggested no inflation of the test statistics due to confounding or other biases, though the trend

Table 5: Replication of Wilmot et al. EWAS (case-control study in school-age)

				Discovery				Replication				
CpG	Gene	Chr	Position	$n_{studies}$	n	Diff.	p	$n_{studies}$	n	B	SE	p
cg05180887	MYT1L	2	1817263	1	92	-0.05	0.04	4	2080	-0.18	0.20	0.43
cg08479516	VIPR2	7	158905536	1	92	-0.03	0.03	4	1900	0.04	0.52	0.57
cg06201514	MYT1L	2	1817409	1	92	-0.08	0.02	5	2295	-0.06	0.11	0.66
cg13444538	VIPR2	7	158905317	1	92	-0.06	0.03	4	2080	-0.06	0.23	0.74
cg10075506	MYT1L	2	1817351	1	92	-0.08	0.04	5	2295	-0.02	0.12	0.88
cg05554000	VIPR2	7	158905015	1	92	-0.05	0.02	5	2295	-0.02	0.16	0.92

Chr Chromosome
$n_{studies}$ Number of studies
n Number of participants
Dff. Difference in methylation between cases and controls (Negative values indicate hypomethylation in controls)
B Regression coefficient
SE Standard error

towards positive associations remained. The pooled results in the meta-analysis had a low lambda (0.96), showed no signs of inflation (0.92) but a slight over-representation of positive associations (0.14).

Single Probe Meta-Analysis

We associated DNA methylation at school-age in whole-blood at 466,574 CpG sites with ADHD symptoms at the same age. No CpG reached genome-wide significance (all p>4.96E-06), see Figure 2 for Manhattan plot. Furthermore, none of the loci whose DNA methylation at birth was significantly associated with ADHD symptoms, also showed a cross-sectional association of DNA methylation at school-age with ADHD symptoms (p>0.33), and 5 out of the 9 regression estimates were in the opposite direction (Table 3).

Replication of previous EWAS

We attempted to replicate six most suggestive EWAS probes of a case-control study, as defined by the authors. While all but one showed a consistent direction in the PACE cohorts, none of the probes were statistically significant. (Table 5)

Stability of methylation association across age

The associations between methylation at birth with ADHD symptoms and methylation at school-age with ADHD symptoms were consistent on the epigenome-wide level. The regression estimates from those CpG sites, which had nominally significant associations at birth (p<0.05, n=73,057) correlated with the regression estimates of the birth EWAS (rs=0.45). When restricting the school-age methylation EWAS to those cohorts, which were not featured in the birth methylation EWAS, the correlation remained (rs=0.30). When filtering for probes which were nominally significant at school-age, 23770 probes remained of which 4075 overlapped with nominally significant probes at birth. The correlation for this set was very similar, rs=0.47 among all cohorts and rs=0.35 between independent cohorts. Thus, regression coefficients based on birth-cord and school-age methylation positively correlate and generalize to independent samples.

DISCUSSION

We performed, in this population-based study, the first epigenome-wide meta-analysis of ADHD symptoms, using two DNA methylation assessments (birth and school-age), as well as repeated measures of ADHD. DNA methylation at birth predicted the later development of ADHD symptoms with a genome-wide significant level at nine loci, but not in school-age. Interestingly all the identified probes showed a pattern of a high

average rate of methylation in cord blood, where lower levels of methylation in an individual were associated with more ADHD symptoms in childhood.

DNA methylation at this stage in life reflects the effects of genetics and the intrauterine environment. The results thus suggest that cord blood DNA methylation is a marker for some of the ADHD risk factors present before birth or a potential mediator of these risk factors. For instance, in utero environmental factors, such as lead exposure, have been associated with ADHD risk. Such influences may be mediated by changes in DNA methylation, which in turn affect gene expression and downstream phenotypes.[34] In this scenario DNA methylation status is involved in the etiology of ADHD and could aid in understanding the psychopathology, as well as give clues to prevention and treatment. While not impossible, reverse causality at this age is unlikely to explain our results, as ADHD only manifests at a later stage of development. However, confounding is a very real possibility. In this example, the lead-altered DNA methylation levels might not affect ADHD risk and simply indicate lead exposure, which may act via different pathways. In this scenario DNA methylation would act as a prognostic marker rather than be on a causal pathway. In this case, DNA methylation status may be useful for prediction and prevention of exposures, but not a treatment target itself.

We analyzed DNA methylation in cord blood, which may not correspond to the methylation status in the brain. Most of the significant probes did not show consistent correlation ($r < 0.1$) between methylation status in whole blood and post-mortem brain tissue in a previous study.[32] However, one probe is the exception: cg01271805 methylation in whole blood is associated with methylation status in various brain regions. Importantly, this probe lies in the promoter region of the gene ERC2, which is highly expressed in brain tissue. ERC2 encodes a protein, which regulates calcium dependent neurotransmitter release in the axonal terminal.[35] Specifically, ERC2 is suspected to increase the sensitivity of voltage dependent calcium channels to hyperpolarization, resulting in higher neurotransmitter release. SNPs in the ERC2 locus were previously used to distinguish schizophrenia and bipolar disorder patients[36] and have been suggested to impact cognitive functioning[37]. ERC2 is especially expressed in Broadmann area 9. Previous imaging studies have demonstrated differential activation in this area when children with or without ADHD performed various cognitive tasks.[38,39] The correlation with brain methylation, the location in a promotor and gene expression in the brain make cg01271805 a plausible candidate locus, where reduced methylation may cuasally affect ADHD development. We hypothesize, that lower methylation levels at cg01271805 increases the expression of ERC2, which in turn increases neurotransmitter release, with an adverse impact on the development of ADHD symptoms. Another gene with a genome-wide signficant probe and high relevance for neural functioning is CREB5 (cg25520701). CREB5 is expressed in fetal brain and the prefrontal cortex, and was previously related to neurite outgrowth. Morever, SNPs in this gene were associat-

ed with ADHD in GWAS.[40,41] Thus it is plausible, that differences in DNA methylation in this locus may modify ADHD risk during developmental stages.

It is noteworthy, that all genome-wide significant CpG sites at birth have high average methylation values and that children with ADHD tended to have lower values. This might indicate a spurious association due to a ceiling effect. However, a closer examination of the distribution reveals that this is less likely. Figure 4 shows a scatterplot of the top CpG sites in the Generation R cohort. While the average values are high, the distribution of the methylation levels is relatively normal, with very few observations at the maximum and there are no children with high ADHD and extremely low methylation values.

While the birth methylation EWAS identified several loci, associating school-age methylation with concurrent ADHD symptoms revealed no genome-wide significant associations and the overall signal was lower, despite similar sample size. None of the probes significantly associated at birth showed any association when measured at school-age. Given that sample sizes were comparable, this difference in predictive ability must come from changes in the epigenome, rather than differences in statistical power. On the one hand, this may be considered surprising given that typically two factors are more strongly associated if measured in closer temporal proximity. On the other hand, Walton et al. observed in a previous EWAS[17], that birth methylation may be a better predictor of later ADHD symptoms than childhood methylation, possibly reflecting sensitive periods. Whether DNA methylation in cord blood has stronger causal effects or is a better marker for early life factors cannot be concluded from the present study. Alternatively, tissue differences between cord blood and whole blood may account for the differences in association pattern. Finally, it is possible that interventions in childhood and other environmental influences diluted the differences in the epigenome between children with more or fewer ADHD symptoms.

That said, we observed a substantial consistency in the associations of methylation at both timepoints with ADHD symptoms. The regression estimates of both EWAS correlated on a genome-wide level. This held true, even if overlapping cohorts were removed from analysis suggesting that the association between DNA methylation at birth and ADHD symptoms to some extent remain in school-age and are consistent across independent cohorts. This implies that DNA methylation in school age may be useful as biomarker for ADHD symptoms, but the development of such a marker would require higher powered studies compared to a biomarker based on cord blood and may be less reliable.

Strengths of this study include the large number of participants and cohorts, the repeated outcome measures, extensive control for potential confounding factors and the use of DNA methylation at two different time-points, enabling to characterize both prospective and cross-sectional associations with ADHD symptoms. However, several limitations need to be discussed as well. A causal interpretation of our findings is challenged by the possibility of residual confounding and reverse causality. For instance,

while we controlled for some potential adverse environments, such as smoking during pregnancy, DNA methylation might be a marker for other adverse environments which could affect ADHD via independent pathways. In addition, children with higher ADHD symptoms may evoke a particular environment, which might shape the epigenome. This is less likely the case for cord blood methylation, but may be a substantial factor for the cross-sectional analyses in school-age. Future studies could explore further causal interpretations of the found associations. It is also likely that many more CpG sites are associated with ADHD than identified in this study. Thus further sample size increases are likely necessary to detect further methylation sites.

In summary, we identified nine CpG sites, in which lower methylation status at birth is associated with later development of ADHD symptoms. The results suggest that DNA methylation in the ERC2 and CREB5 gene may exert an influence on ADHD symptoms, potentially via modification of neurotransmitter functioning or neurite outgrowth. None of the sites prospectively associated with ADHD in cord blood were cross-sectionally associated with ADHD when measured during school-age, and generally no genome-wide significant CpGs were identified in childhood. However, on an epigenome-wide level the association of the methylation probes with ADHD showed consistency across both time-points, thus development of biomarkers which are predictive of ADHD at any age may be possible.

REFERENCES

1. Thomas R, Sanders S, Doust J, Beller E, Glasziou P. Prevalence of attention-deficit/hyperactivity disorder: a systematic review and meta-analysis. Pediatrics 2015; 135: e994-1001.
2. Bergen SE, Gardner CO, Kendler KS. Age-related changes in heritability of behavioral phenotypes over adolescence and young adulthood. Twin Res Hum Genet 2007; 10: 423–433.
3. Larsson H, Chang Z, D'Onofrio BM, Lichtenstein P. The heritability of clinically diagnosed attention deficit hyperactivity disorder across the lifespan. Psychol Med 2014; 44: 2223–2229.
4. Swanson JM, Kinsbourne M, Nigg J, Lanphear B, Stefanatos GA, Volkow N et al. Etiologic Subtypes of Attention-Deficit/Hyperactivity Disorder: Brain Imaging, Molecular Genetic and Environmental Factors and the Dopamine Hypothesis. Neuropsychol Rev 2007; 17: 39–59.
5. Marceau K, Cinnamon Bidwell L, Karoly HC, Evans AS, Todorov AA, Palmer RH et al. Within-Family Effects of Smoking during Pregnancy on ADHD: the Importance of Phenotype. J Abnorm Child Psychol 2018; 46: 685–699.
6. Daneshparvar M, Mostafavi SA, Jeddi MZ, Yunesian M, Mesdaghinia A, Mahvi AH et al. The role of lead exposure on Attention-Deficit/Hyperactivity Disorder in children: A systematic review. Iran J Psychiatry 2016; 11: 1–14.
7. Lam J, Lanphear BP, Bellinger D, Axelrad DA, McPartland J, Sutton P et al. Developmental pbde exposure and IQ/ADHD in childhood: A systematic review and meta-analysis. Environ Health Perspect 2017; 125. doi:10.1289/EHP1632.
8. Barker ED, Walton E, Cecil CAM. Annual Research Review: DNA methylation as a mediator in the association between risk exposure and child and adolescent psychopathology. J Child Psychol Psychiatry 2017; 4: 303–322.
9. Dall'Aglio L, Muka T, Cecil CAM, Bramer WM, Verbiest MMPJ, Nano J et al. The Role of Epigenetic Modifications in Neurodevelopmental Disorders: A Systematic Review. Neurosci Biobehav Rev 2018; 94: 17–30.
10. Pappa I, Mileva-Seitz VR, Szekely E, Verhulst FC, Bakermans-Kranenburg MJ, Jaddoe VWV et al. DRD4 VNTRs, observed stranger fear in preschoolers and later ADHD symptoms. Psychiatry Res 2014; 220: 982–986.
11. Adriani W, Romano E, Pucci M, Pascale E, Cerniglia L, Cimino S et al. Potential for diagnosis versus therapy monitoring of attention deficit hyperactivity disorder: a new epigenetic biomarker interacting with both genotype and auto-immunity. Eur Child Adolesc Psychiatry 2018; 27: 241–252.
12. van Mil NH, Steegers-Theunissen RPM, Bouwland-Both MI, Verbiest MMPJ, Rijlaarsdam J, Hofman A et al. DNA methylation profiles at birth and child ADHD symptoms. J Psychiatr Res 2014; 49: 51–9.
13. Xu Y, Chen X-T, Luo M, Tang Y, Zhang G, Wu D et al. Multiple epigenetic factors predict the attention deficit/hyperactivity disorder among the Chinese Han children. J Psychiatr Res 2015; 64: 40–50.
14. Dadds MR, Schollar-Root O, Lenroot R, Moul C, Hawes DJ. Epigenetic regulation of the DRD4 gene and dimensions of attention-deficit/hyperactivity disorder in children. Eur Child Adolesc Psychiatry 2016; 25: 1081–1089.
15. Ding K, Yang J, Reynolds GP, Chen B, Shao J, Liu R et al. DAT1 methylation is associated with methylphenidate response on oppositional and hyperactive-impulsive symptoms in children and adolescents with ADHD. World J Biol Psychiatry 2017; 18: 291–299.

16. Wilmot B, Fry R, Smeester L, Musser ED, Mill J, Nigg JT. Methylomic analysis of salivary DNA in childhood ADHD identifies altered DNA methylation in VIPR2. J Child Psychol Psychiatry Allied Discip 2016; 57: 152–160.
17. Walton E, Pingault J-B, Cecil CAM, Gaunt TR, Relton CL, Mill J et al. Epigenetic profiling of ADHD symptoms trajectories: a prospective, methylome-wide study. Mol Psychiatry 2017; 22: 250–256.
18. 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 2018; 47: 22–23u.
19. Maarten van Iterson, Erik van Zwet , the BIOS Consortium PES and BT, Heijmans. Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. bioRxiv 2016; 38: 1–16.
20. Bakulski KM, Feinberg JI, Andrews S V., Yang J, Brown S, L. McKenney S et al. DNA methylation of cord blood cell types: Applications for mixed cell birth studies. Epigenetics 2016; 11: 354–362.
21. R Core Team. R: A Language and Environment for Statistical Computing. 2016.https://www.r-project.org/.
22. Bates D, Mächler M, Bolker BM, Walker SC. Fitting Linear Mixed-Effects Models using lme4. J Stat Softw 2014; : 1–51.
23. Han B, Eskin E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. Am J Hum Genet 2011; 88: 586–598.
24. Smith AK, Kilaru V, Kocak M, Almli LM, Mercer KB, Ressler KJ et al. Methylation quantitative trait loci (meQTLs) are consistently detected across ancestry, developmental stage, and tissue type. BMC Genomics 2014; 15. doi:10.1186/1471-2164-15-145.
25. Hannon E, Knox O, Sugden K, Burrage J, Wong CCY, Belsky DW et al. Characterizing genetic and environmental influences on variable DNA methylation using monozygotic and dizygotic twins. PLoS Genet 2018; 14: 1–27.
26. Phipson B, Maksimovic J, Oshlack A. missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. Bioinformatics 2015; 32: 286–288.
27. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics 2011; 27: 1739–1740.
28. Geeleher P, Hartnett L, Egan LJ, Golden A, Raja Ali RA, Seoighe C. Gene-set analysis is severely biased when applied to genome-wide methylation data. Bioinformatics 2013; 29: 1851–1857.
29. IlluminaHumanMethylation450kanno HK. ilmn12. hg19: Annotation for Illumina's 450k methylation arrays. R Packag version 02 2014; 1.
30. Maitre L, de Bont J, Casas M, Robinson O, Aasvang GM, Agier L et al. Human Early Life Exposome (HELIX) study: a European population-based exposome cohort. BMJ Open 2018; 8: e021311.
31. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics 2012; 13. doi:10.1186/1471-2105-13-86.
32. Edgar RD, Jones MJ, Meaney MJ, Turecki G, Kobor MS. BECon: A tool for interpreting DNA methylation findings from blood in the context of brain. Transl Psychiatry 2017; 7: e1187-10.
33. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S et al. The genotype-tissue expression (GTEx) project. Nat Genet 2013; 45: 580–585.

34. Wu S, Hivert MF, Cardenas A, Zhong J, Rifas-Shiman SL, Agha G et al. Exposure to low levels of lead in utero and umbilical cord blood DNA methylation in project viva: An epigenome-wide association study. Environ Health Perspect 2017; 125: 1–10.
35. Kiyonaka S, Nakajima H, Takada Y, Hida Y, Yoshioka T, Hagiwara A et al. Physical and functional interaction of the active zone protein CAST/ERC2 and the β-subunit of the voltage-dependent Ca2+ channel. J Biochem 2012; 152: 149–159.
36. Curtis D, Vine AE, McQuillin A, Bass NJ, Pereira A, Kandaswamy R et al. Case-case genome-wide association analysis shows markers differentially associated with schizophrenia and bipolar disorder and implicates calcium channel genes. Psychiatr Genet 2011; 21: 1–4.
37. Hatzimanolis A, Bhatnagar P, Moes A, Wang R, Roussos P, Bitsios P et al. Common genetic variation and schizophrenia polygenic risk influence neurocognitive performance in young adulthood. Am J Med Genet Part B Neuropsychiatr Genet 2015; 168: 392–401.
38. Epstein JN, Delbello MP, Adler CM, Altaye M, Kramer M, Mills NP et al. Differential patterns of brain activation over time in adolescents with and without attention deficit hyperactivity disorder (ADHD) during performance of a sustained attention task. Neuropediatrics 2009; 40: 1–5.
39. Monden Y, Dan H, Nagashima M, Dan I, Tsuzuki D, Kyutoku Y et al. Right prefrontal activation as a neuro-functional biomarker for monitoring acute effects of methylphenidate in ADHD children: An fNIRS study. NeuroImage Clin 2012; 1: 131–140.
40. Franke B, Neale BM, Faraone S V. Genome-wide association studies in ADHD. Hum Genet 2009; 126: 13–50.
41. Klein M, Walters RK, Demontis D, Stein JL, Hibar DP, Adams HH et al. Genetic markers of ADHD-related variations in intracranial volume. Am J Psychiatry 2019; 176: 228–238.