Epigenome-wide association study of ADHD symptoms in adults


PII: S0006-3223(19)30125-8
DOI: https://doi.org/10.1016/j.biopsych.2019.02.016
Reference: BPS 13775

To appear in: Biological Psychiatry

Received Date: 8 July 2018
Revised Date: 7 February 2019
Accepted Date: 8 February 2019


This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
Epigenome-wide association study of ADHD symptoms in adults

Jenny van Dongen¹, Nuno R. Zilhão¹, Karen Sugden²,³, BIOS consortium⁴, Eilis J. Hannon⁵, Jonathan Mill⁵, Avshalom Caspi²,³,⁶, Jessica Agnew-Blais⁶, Louise Arseneault⁶, David L. Corcoran⁴, Terrie E. Moffitt²,³,⁶, Richie Poulton⁷, Barbara Franke⁸, Dorret I. Boomsma¹

Running title: Epigenome-wide association study of ADHD symptoms

¹Department of Biological Psychology, Amsterdam Public Health research institute, Vrije Universiteit Amsterdam, Van der Boechorststraat 7-9, 1081BT, Amsterdam, The Netherlands
²Department of Psychology and Neuroscience, Duke University, Durham, NC, USA
³Center for Genomic and Computational Biology, Duke University, Durham, NC, USA
⁴Biobank-based Integrated Omics Studies Consortium. For a complete list of authors, see the acknowledgements.
⁵University of Exeter Medical School, University of Exeter, Exeter, UK
⁶Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King’s College London, London, UK
⁷Dunedin Multidisciplinary Health and Development Research Unit, Department of Psychology, University of Otago, Dunedin, New Zealand
⁸Departments of Human Genetics and Psychiatry, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, The Netherlands

Corresponding author:
Jenny van Dongen
Department of Biological Psychology
Vrije Universiteit Amsterdam
Van der Boechorststraat 7-9
1081 BT Amsterdam, The Netherlands
T: +31 (0) 20 598 3570
F: +31-(0) 20 598 8832
E: j.van.dongen@vu.nl

Key words: ADHD, meta-analysis, DNA methylation, EWAS, CAARS, epigenetic

Number of words: abstract; 250. Main text; 3998

Number of Tables: 3, Figures: 1, Supplemental files: 2 (1 PDF with text and figures; 1 Excel with Tables).
Abstract

Background

Previous studies have reported associations between ADHD symptoms and DNA methylation in children. We report the first epigenome-wide association study (EWAS) meta-analysis of adult ADHD symptoms, based on peripheral blood DNA methylation (HM450 array) in three population-based adult cohorts.

Methods

An EWAS was performed in the Netherlands Twin Register (NTR, N=2258, mean age 37 years), the Dunedin Study (N=800, age 38 years), and the Environmental Risk Longitudinal Twin Study (E-risk, N=1631, age 18 years), and results were combined through meta-analysis (total sample size= 4689). Second, region-based analyses accounting for the correlation between nearby methylation sites were performed.

Results

One epigenome-wide significant differentially methylated position (DMP) was detected in the Dunedin Study, but meta-analysis did not detect DMPs that were robustly associated across cohorts. In region-based analyses, six significant differentially methylation regions (DMRs) were identified in NTR, 19 in the Dunedin Study and none in E-risk. Of these DMRs, 92% were associated with methylation QTLs, and 68% showed moderate to large blood-brain correlations for DNA methylation levels. DMRs included six non-overlapping DMRs (three in NTR, three in Dunedin) in the Major Histocompatibility Complex (MHC), which were associated with expression of genes in the MHC, including C4A and C4B, previously implicated in schizophrenia.

Conclusions

Our findings point at new candidate loci involved in immune and neuronal functions that await further replication. Our work also illustrates the need for further research to examine to what extent epigenetic associations with psychiatric traits depend on characteristics such as age, comorbidities, exposures, and genetic background.
Introduction

Attention-deficit/hyperactivity disorder (ADHD) is among the top-ranking psychiatric diagnoses in children and adults(1), and appears to reflect the extreme end of a continuous distribution of ADHD symptoms in the population(2–4). Genetic association studies have identified common and rare variants associated with ADHD(5–7). Numerous environmental risk factors have been reported, including pre- and perinatal factors(8) (birth weight(9), prenatal exposure to maternal smoking(10), toxins(8), maternal fever and infections(11)) and postnatal factors (childhood maltreatment(12), current stress(12), and infections(13)). Recent studies have investigated if DNA methylation levels, which regulate gene expression, are associated with ADHD(14–26). DNA methylation might represent a marker that captures the cumulative effects of genetic variants, stochastic effects, and environmental exposures(27, 28) associated with a trait. Several prenatal exposures, including maternal famine(29), maternal folate(30), and maternal smoking(31), and later life exposures, for example, smoking(32), have been associated with stable long-term changes in DNA methylation in blood and other peripheral tissues. Importantly, DNA methylation patterns are largely tissue-specific. ADHD symptom level-associated DNA methylation differences in peripheral tissues such as blood are likely to reveal epigenetic consequences of differential life conditions that correlate with ADHD (biomarker of exposures). The extent to which DNA methylation in peripheral tissues is informative about epigenetic mechanisms that contribute to inter-individual differences in ADHD symptoms or that correlate with causal epigenetic mechanisms in other tissues is unknown. It has been reported that DNA methylation levels in blood correlate to a limited extent with methylation levels in other tissues, including the brain(33–36). One explanation for such correlations is that methylation Quantitative Trait Locus (mQTL) effects correlate to some extent across tissues, as was recently demonstrated for cis mQTL effects in blood and brain(37).
Several candidate gene studies, most of limited sample size (mean=192, range=82-426), have reported associations between ADHD symptoms in children and DNA methylation in cord blood(38), peripheral blood(19, 21, 39), buccal samples(23), and saliva(21). Other candidate approaches in small samples have reported relationships between DNA methylation, ADHD symptoms, and environmental risk factors(16, 22, 24). To date, three epigenome-wide association studies (EWASs) of ADHD symptoms in children have been published(20, 25, 40). The first study measured DNA methylation in saliva in a small group of boys with ADHD and controls (age 7-12, sample size=112), and reported genes with suggestive evidence for association(20). The second study measured DNA methylation in 384 cord blood samples and found no significant differences between children later diagnosed with ADHD and controls(41). The largest EWAS (817 children from a UK population-based study) identified 13 loci, where DNA methylation level in cord blood was significantly predictive of ADHD symptom trajectories between age 7 and 15 years(40); among those was ST3GAL3; one of the significant loci from the Genome-wide Association Study (GWAS) of ADHD(7). Since DNA methylation and ADHD symptoms both change with age, it is unknown if findings in children persist in adulthood. Moreover, the robustness of associations between ADHD symptoms and DNA methylation remains to be investigated in larger studies of multiple cohorts.

Here, we report the first EWAS meta-analysis of ADHD symptoms in adults; this is also the largest EWAS of ADHD symptoms to date. We assessed the association between whole blood DNA methylation and ADHD symptoms in three population-based cohorts: the Netherlands Twin Register (NTR)(42), the Dunedin Multidisciplinary Health and Development Study from New Zealand, and the Environmental risk (E-risk) twin study from the UK. In secondary analyses, we 1) tested if CpGs with a lower, albeit non-significant, p-value showed enrichment for loci previously identified in GWASs of psychiatric disease or in EWASs of psychiatric phenotypes or exposures, 2) analysed differentially methylated regions (DMRs), to examine the evidence for small methylation differences at multiple
nearby CpGs, and 3) examined the relationship between DNA methylation and RNA transcript levels in blood, effects of mQTLs, and the correlation between DNA methylation in blood and brain, to facilitate the biological interpretation of findings.
Methods and Materials

Overview

The EWAS was performed in three cohorts (Figure 1): NTR(42) (N=2232 individuals from twin families), the Dunedin study(43) (N=800 unrelated individuals) and E-risk(44) (N=1631 twins). Results were combined in a meta-analysis to identify DMPs and to examine heterogeneity of effects across cohorts. Next, we performed secondary analyses. Genome-wide meta-analysis test statistics were used to test for enrichment of ADHD EWAS signals in loci detected in previous GWA and EWA studies of relevant traits. We performed DMR analyses in each cohort, compared results across cohorts, and performed a meta-analysis-based DMR analysis. Functional follow-up analyses were performed on top-DMPs from the meta-analysis (nominal p-value<1.0x10^{-5}) and on significant DMRs. Further details about the analyses are provided in Appendix 1.

Ethical permission

Ethical permission is described in Appendix 1. Written informed consent was obtained from all participants.

ADHD symptoms

The EWAS in NTR was performed on the Conners Adult ADHD Rating Scale (CAARS) index (total ADHD symptoms; Figure S1a)(45). Sensitivity analyses were performed on DSM-IV-based CAARS inattention and hyperactivity subscales, available for 1846 samples. In the Dunedin Study, DSM-5 ADHD symptoms (Figure S1b) were assessed based on private structured interviews as described previously(46) at age 38 years. In the E-Risk Study, DSM-5 ADHD symptoms (Figure S1c) were assessed based on private structured interviews as described previously(47) at age 18 years.
Peripheral blood DNA methylation

DNA methylation was assessed with the Infinium HumanMethylation450 BeadChip Kit (HM450k; Illumina, San Diego, CA, USA). Normalization was performed with functional normalization in NTR(48), with methylumi in the Dunedin Study, and with dasen in E-Risk(49). The following probes were removed from all cohorts: sex chromosomes, probes with a single nucleotide polymorphism (SNP) within the CpG site (at the C or G position), irrespective of minor allele frequency (MAF), based on the genome of the Netherlands (GONL) reference population(50), probes with common (>5% MAF) SNPs within 10 base pairs (bp) of the single base extension site(51), and ambiguous mapping probes with an overlap of at least 47 bases per probe(52). Only methylation sites that were present in all cohorts were kept in the analysis, leaving 394194 sites.

Statistical analyses

Epigenome-wide association study (EWAS)

The association between DNA methylation levels and ADHD symptoms was tested for each site under a linear model (Dunedin) or generalized estimation equation (gee) model accounting for relatedness of twins and other family members (NTR and E-risk); methylation β-value were assessed as outcome and the following predictors were used: ADHD symptoms, sex, smoking status, white blood cell (WBC) percentages, age at blood sampling (only in NTR), cohort-specific technical covariates (i.e. sample plate and array row or Principal Components (PCs) based on control probes), and Principal Components (PCs) based on genome-wide SNPs. The R-package bacon was used to compute the Bayesian inflation factor(53).
Meta-analysis

A p-value-based fixed-effects sample size-weighted meta-analysis was performed in METAL. The sample size-weighted method was chosen because of the differences in measurement scales of ADHD symptoms across studies. Statistical significance was assessed considering Bonferroni correction for the number of methylation sites tested (alpha=0.05/394194=1.3x10^{-7}). False discovery rate (FDR) q-values are presented in supplementary tables. The I^2 statistic provided by METAL was evaluated to assess heterogeneity.

Inattention and hyperactivity

For top-DMPs, sensitivity analyses were performed in which the association of these DMPs with inattention and hyperactivity/impulsivity subscales was tested in NTR.

Overlap with EWAS and GWAS loci

Enrichment analysis was performed to examine whether CpGs in or near loci detected by GWASs, or top-ranking CpGs from previous EWA studies, on average, showed a stronger association with ADHD symptoms than other genome-wide HM450 methylation sites. We considered the most recent GWASs for ADHD(7), major depressive disorder (MDD)(54), schizophrenia(55), and autism spectrum disorders (ASD)(56), and the largest available EWA studies on ADHD symptoms in children(40), schizophrenia(57), smoking(32), and maternal smoking(31).

Differentially methylated regions

We used the python module ‘Comb-p’(58) to identify regions where multiple correlated methylation sites show evidence for association with ADHD symptoms. We report significant regions (Šidák p<0.05) with at least two methylation sites within a 500bp window. Comb-p was applied in each of the 3 cohorts, separately, and on the meta-analysis results.
Functional follow-up analyses

In follow-up analyses, previously published datasets were used to test if DNA methylation level at top-DMPs and at CpGs within significant DMRs were associated with whole blood gene expression level in \textit{cis}(59), whole blood mQTLs(59), and to examine the correlation between DNA methylation in blood and four brain regions (prefrontal cortex, entorhinal cortex, superior temporal gyrus and cerebellum)(35).

Power analysis

Power analyses are described in Appendix 1.
Results

Meta-analysis

Demographic information of the cohorts is provided in Table 1. Genome-wide EWAS test statistics from each cohort showed no inflation (Table S1, Figure S2). One significant DMP (alpha=1.3x10^{-7}) was detected in Dunedin (cg26197679, chromosome 8 intergenic), and none in NTR and E-risk. Meta-analysis of the three cohorts (N=4689) detected no significant DMPs, and meta-analysis test statistics showed no inflation (Table S1, Figure S2). Summary statistics for genome-wide methylation sites are provided in Table S2. Top-DMPs with a nominal p-value (p<1.0x10^{-5}) are presented in Table 2. Two DMPs showed a negative relationship with ADHD symptoms (cg26197679, intergenic and cg23144852, gene body OPA1), and one DMP showed a positive relationship with ADHD symptoms in all cohorts (cg10984962, gene body AGAP1). The effect sizes for these DMPs in the individual cohorts are presented in Table S3. In NTR, an increase of one standard deviation in CAARS ADHD symptoms (3.9 points) was associated with a methylation change of -0.04%(cg26197679), -0.16%(cg23144852), and 0.10%( cg10984962), respectively.

Heterogeneity of top-DMPs

One top-DMP (cg26197679) displayed large between-study heterogeneity (I^2=93.6%, heterogeneity p=1.6x10^{-7}). This DMP showed the strongest association with ADHD symptoms and was epigenome-wide significant in the Dunedin study, where an increase of one standard deviation in DSM-5 ADHD symptoms was associated with a methylation change of -0.62%. Although the effect size was weaker in the other cohorts, the direction of association was the same in all cohorts (Table 2). Inspection of all top-ranking sites from the individual cohorts (p<1.0x10^{-5}: 14 sites) revealed that all sites were characterized by substantial between-study heterogeneity (mean I^2=87.2%, range=71.3%-93.8%, Table S4). Nine sites (64.3%) showed the same direction of effect across all cohorts.
Inattention and hyperactivity

In sensitivity analyses, the association with inattention and hyperactivity/impulsivity subscales, separately, was tested in NTR for the three top-DMPs from the meta-analysis. This analysis showed that the direction of effect and the strength of the association with each of the subscales was highly similar (Table S5).

Enrichment of EWAS and GWAS loci

In testing for overlap of our EWAS meta-analysis results with findings from previous EWASs and GWASs, we failed to observe enrichment of CpGs previously associated with longitudinal ADHD trajectories in children(40) or schizophrenia in adults(57), CpGs associated with individual(32) or maternal smoking(31), and CpGs near GWAS loci for ADHD, ASD, MDD, or schizophrenia (Table S6). Methylation sites previously associated with schizophrenia(57) showed a small but significant depletion of signal for ADHD symptoms.

Differentially methylated regions

In NTR, six significant DMRs were identified, which spanned from 164 bp to 848 bp, and included three to 32 CpGs (Table 3 and Table S7). One example is illustrated in Figure S3. In the Dunedin study, 19 significant DMRs were identified, spanning from two to 51 CpGs within regions of 18 bp to 1818 bp (Table 3 and Table S7). In E-risk, no significant DMRs were identified. Six distinct DMRs were detected in the Major Histocompatibility Complex (MHC) region (chromosome 6): three in NTR, and three in the Dunedin Study. In line with the heterogeneity of DMP results across cohorts, none of the DMRs detected in NTR and Dunedin overlapped, and DMR analysis on the meta-analysis of the three cohorts did not detect significant DMRs.
Significant DMRs did not overlap with CpGs from the previous EWAS of ADHD trajectories in children (17) or with schizophrenia in adults (57). One of the six DMRs identified in NTR (chromosome 6: 33245460-3324630) contained CpGs previously associated with smoking (12 of the 32 CpGs) and maternal smoking (3 CpGs). Five of the 19 DMRs identified in the Dunedin Study contained CpGs associated with smoking or maternal smoking (Table S7).

Several DMRs were located in proximity of SNPs associated with schizophrenia (55): of the DMRs in NTR, two were located within 100 kb and one was located within 1 Mb of schizophrenia-associated SNPs, respectively (all in the MHC region). Five DMRs in the Dunedin study were located within 1 Mb of schizophrenia-associated SNPs (two on chromosome 6, two on chromosome 7, and one on chromosome 15). None of the DMRs was located within 1 Mb of significant GWAS loci for ADHD, MDD, or ASD.

**Gene expression in cis**

To examine potential functional consequences top-DMPs and DMRs, we used previously published data on whole blood DNA methylation and RNA-sequencing (n=2,101 samples). While DNA methylation levels at the ADHD symptom level top-DMPs were not associated with RNA levels of genes in cis, methylation level at CpGs within five of the six significant DMRs detected in NTR were associated with expression levels of 14 genes (Table 3 and Table S8): At one DMR, higher methylation level correlated with lower expression, at another DMR, higher methylation level correlated with higher expression, and at three DMRs, expression of some genes correlated positively and others negatively with methylation level. Of the 19 DMRs identified in the Dunedin Longitudinal Study, the methylation levels at CpGs within three DMRs were associated expression levels of seven genes (Table 3 and Table S9): at two DMRs, a higher methylation level correlated with lower expression, and at one DMR, a higher methylation level correlated with higher expression.

**mQTLs**
To gain insight into genetic causes of variation underlying top-DMPs and DMRs, we obtained whole blood mQTL data (n=3841 samples) (59). One of the three top-DMPs from the ADHD symptom meta-analysis was associated with six independent SNPs (mQTLs) in cis (Table S10). The majority of DMRs (92.0%; 23) was associated with mQTLs. For ADHD-associated DMRs in NTR, on average 68% of the CpGs within a DMR (range=36.4-92.3%) was associated with at least one mQTL SNP. In total, 164 mQTL associations were identified for NTR DMRs (76.8% were cis mQTLs and 23.2% were trans mQTLs; Table S11), involving 59 CpGs and 55 SNPs. For ADHD-associated DMRs in Dunedin, 323 mQTL associations were identified, involving 126 CpGs and 154 SNPs (88% cis and 13% trans; Table S12). On average, 64.9% (range=0-100%) of CpGs within DMRs identified in the Dunedin study was associated with one or more mQTLs. We highlight one example; a DMR on chromosome 11, detected in NTR and associated with the expression of ACY3, was associated with cis and trans mQTLs, and the correlation structure of DNA methylation within this DMR (Figure S2) mirrored the sharing across CpGs of trans-mQTLs on chromosome 6 (Figure S3). Comparing the overlap of mQTLs across cohorts revealed that two DMRs were associated with a common set of trans mQTLs. The DMR on chromosome 11 in NTR (associated with expression of ACY3), and the DMR on chromosome 15 in Dunedin (associated with expression of SEMA4B) were both associated in trans with SNPs on chromosome 6:109626965-109616420 (rs9374080, rs1008084, and rs9386791).

**Correlation between DNA methylation level in blood and brain**

While DNA methylation levels at the ADHD symptom top-DMPs did not correlate significantly between blood and brain, CpGs within 4 of the 6 DMRs (66.7%) detected in NTR (Table S13), and 13 of the 19 DMRs (68.4%) detected in the Dunedin Study (Table S14) showed significantly correlated DNA methylation levels between blood and one or multiple brain regions. For all CpGs, except for one, the correlation was positive (mean r=0.50, range=-0.42-0.70). The number of CpGs per DMR that showed
correlated methylation levels between blood and brain ranged from 1 to 31 (mean=5.5). The DMR with the largest number of CpGs with significant blood-brain correlations (31 CpGs), was a DMR in the MHC region, detected in the Dunedin study and associated with the expression level of multiple genes, including C4A and C4B. An exemplary plot of methylation levels in blood and brain for one CpG in this region (cg01337207) is provided in Figure S5.
Discussion

We performed an EWAS of ADHD symptoms in three population-based adult cohorts. Our hypothesis was that DNA methylation in blood may provide insight into epigenetic consequences of life conditions that correlate with ADHD symptoms, and potentially into epigenetic mechanisms that contribute to ADHD symptoms or that correlate with causal epigenetic mechanisms in the brain. In the Dunedin Study, we identified one significant DMP, where a higher methylation level correlated with fewer ADHD symptoms (cg26197679, chr8 intergenic). This CpG, as well as other top-ranking CpGs from individual cohorts showed considerable heterogeneity, and no significant DMPs were detected in a meta-analysis of the three cohorts. In secondary, region-based analyses, we tested if ADHD symptoms were associated with methylation differences at multiple nearby CpGs that individually failed to reach epigenome-wide significance. We identified six significant DMRs in NTR, 19 in the Dunedin study, and none in E-risk. In line with the heterogeneity of DMP results, none of the DMRs overlapped across cohorts, and the meta-analysis-based DMR analysis did not detect significant DMRs. With respect to effects of differential exposures, although the EWAS signal showed no significant enrichment for CpGs previously associated with smoking, some DMRs contained CpGs previously associated with smoking(32) or maternal smoking(31), even after we adjusted for smoking. There are several possible explanations for this finding(60): 1) residual confounding effects of smoking, 2) second-hand smoking exposure, including maternal prenatal smoking, 3) exposures other than smoking; for instance, ADHD is associated with general substance use(61), and 4) these CpGs are connected to a shared underlying biology of ADHD symptoms and smoking. Importantly, the current study shows that effects of mQTLs and correlated methylation levels between blood and brain exceed effects of smoking: 92% of DMRs was associated with genetic variants, 68% of DMRs showed correlated methylation levels between blood and brain regions, and 24% of DMRs were associated with smoking. These observations suggest that inter-individual differences in DNA methylation at these DMRs are not merely driven by lifestyle differences.
associated with ADHD symptoms such as smoking, and that some of the methylation differences in whole blood associated with ADHD symptoms might be a marker for methylation variation in the brain.

Some top-DMPs and top-DMRs mapped to genes that have been previously linked to psychiatric disorders or implicated in brain biology. These are potential candidates for being involved in the underlying biology of ADHD symptoms, provided that these loci also show symptom-associated differences in epigenetic regulation in the brain. For instance, cg10984962; the third-ranking CpG from the meta-analysis, is located in AGAPI, which encodes a protein involved in endosomal trafficking. In neuronal cells, it plays a role in the recycling of muscarinic acetylcholine receptors(62) and was shown to influence dendritic spine morphology(63). Yet, methylation levels in blood did not correlate with methylation levels in the brain at this CpG. Six significant DMRs in distinct sub-regions of the MHC were identified in NTR and the Dunedin Study. The top-DMR in MHC in the Dunedin Study was associated with expression levels of SKIV2L, C4B, C4A, TNXB, and TNXA. CpGs in this DMR were not associated with smoking(32). The C4 genes are of great interest, as they have been previously implicated in functional effects of schizophrenia-associated SNPs in the MHC on postnatal synaptic pruning(64). At many CpGs in this DMR, DNA methylation level in blood showed moderate to strong correlations with DNA methylation levels in multiple brain regions. Yet, since the DMRs did not replicate across cohorts, it remains to be established whether they are relevant to ADHD symptom levels.

The lack of overlap of our findings with CpGs from a previous study, at which methylation level in cord blood significantly predicted ADHD trajectories in childhood(40), could indicate that epigenetic associations relevant to ADHD symptoms are age-specific (which has already been described for genetic contributions to ADHD symptoms(65)). The loci detected in cord blood also did not show association with ADHD symptoms when methylation was assessed at age seven in the previous study(40). Age-
specific epigenetic associations may also potentially explain a lack of overlap with GWAS loci. While our EWAS only included adults, the GWAS of ADHD contained mainly children.[7]

This is the first EWAS of ADHD symptoms in adults, the largest epigenetic study of ADHD symptoms to date, and the first study of ADHD symptoms to apply a multi-cohort approach. This study also has limitations. Although all cohorts included adults and applied continuous measures of ADHD symptoms reported by the same informant (self-report), there were also differences between cohorts that may have reduced power. First, the CAARS was used in NTR, which is based on DSM-IV, while DSM-5 symptoms were assessed in the E-risk and Dunedin Study. Second, ADHD symptoms were assessed by a self-report scale in NTR and by structured interviews in the Dunedin Study and E-risk. Third, DNA methylation measurements, processing of the data, and quality control were performed separately in each cohort, based on QC and normalization pipelines that were optimized for each cohort. Fourth, cohorts varied in age and reported number of ADHD symptoms. NTR included a broad age range (mean age=37), while participants from the E-risk study were young adults (age=18), and the age of participants from the Dunedin Study was 38. The prevalence of ADHD in these cohorts, based on the instruments used in the current study, has been previously reported: 7% in NTR(66), 3% in the Dunedin Study(46), and 8% in E-risk (47). Fifth, the cohorts were from different countries, and it is possible that epigenetic differences exist between cohorts due to differences in genetic background or in the presence and frequency of environmental exposures.

We performed a power analysis for the three meta-analysis top-DMPs, based on their effect size observed in NTR (Figure S6). The effect sizes of these three DMPs ranged from 0.09%- 0.92% explained variance. For the top-DMP that showed the least heterogeneity (cg10984962; AGAPI; variance explained= 0.28% in NTR), the required sample size to achieve 80% power is 13508. Similar to our
population-based EWAS of ADHD symptoms, a recent blood-based EWAS meta-analysis of cognitive abilities identified only a few significant sites for most cognitive measures in >6000 individuals(67).

Several limitations apply to epigenetic epidemiology studies including ours. Firstly, the Illumina 450k array captures only ~1.7% of all CpGs in the genome, and the removal of methylation probes that overlap with genetic polymorphisms, albeit inevitable because these polymorphisms compromise the quality of microarray-based measurements of DNA methylation, may limit the ability to detect DNA methylation-mediated SNP effects on the phenotype. Secondly, blood is unlikely to provide a complete picture of ADHD-related epigenetic processes, because epigenetic marks are largely tissue-specific. Yet, methylation changes in blood have already been associated with ADHD symptoms(40), other psychiatric conditions(57), cognition(67), and hippocampal volume(68) in earlier studies. Furthermore, we found that several of the DMRs detected in blood contain CpGs with correlated methylation levels across blood and brain. Finally, a general constraint in epigenetic studies is that trait-associated variation in DNA methylation may arise secondary to trait development (reverse causality) or may be a marker of trait-associated environmental exposures (such as lifestyle, medication use, infections, and early life environmental factors, such as nutrition, and maternal smoking). These associations nevertheless may provide valuable insight into the underlying biological changes associated with ADHD and its risk factors. The (direction of) causality may be addressed in human studies, for example, with longitudinal study designs or Mendelian Randomization(69).

In conclusion, by performing an EWAS of ADHD symptoms in three population-based adult cohorts, we found no significant sites in an EWAS meta-analysis of 4689 individuals and observed considerable heterogeneity of effects across cohorts. We found several significant cohort-specific DMPs and DMRs, with the MHC region emerging in two cohorts. The significance of these findings is unknown, but they may point at new candidate pathways awaiting further replication. We conclude that larger
studies are necessary to identify methylation sites in whole blood that are robustly associated with population-based ADHD symptoms in adults. Our findings also illustrate the need for further research to examine to what extent epigenetic associations with psychiatric traits depend on characteristics such as age, sex, lifetime exposures, genetic background, symptom severity, and comorbidities.
Acknowledgements

Netherlands Twin Register: We would like to thank the twins and their family members for their participation. This study was funded by: BBRMI-NL (NWO 184.021.007), NWO 480-15-001/674; Netherlands Twin Registry Repository: researching the interplay between genome and environment, and ACTION. ACTION receives funding from the European Union Seventh Framework Program (FP7/2007-2013) under grant agreement no 602768. Prof. Boomsma gratefully acknowledges the Royal Netherlands Academy of Science Professor Award (PAH/6635). Prof. Franke receives support from a personal grant from the Netherlands Organization for Scientific Research (NWO; Vici grant 016-130-669), from a grant for the Dutch National Science Agenda for the NWANeurolabNL project (grant 400 17 602), from the European Community’s FP7 under grant agreement no 602805 (Aggressotype), and from the European Community’s Horizon 2020 Program (H2020/2014 – 2020) under grant agreement no 728018 (Eat2beNICE).

Dunedin Study: We thank the Dunedin Study members, Unit research staff, and Study founder Phil Silva. The Dunedin Multidisciplinary Health and Development Research Unit is supported by the New Zealand Health Research Council and New Zealand Ministry of Business, Innovation and Employment (MBIE). This research received support from the US-National Institute of Aging (grant number R01AG032282) and the UK Medical Research Council (grant number MR/P005918). Additional support was provided by the Jacobs Foundation. This work used a high-performance computing facility partially supported by grant 2016-IDG-1013 (“HARDAC+: Reproducible HPC for Next-generation Genomics”) from the North Carolina Biotechnology Center.

E-risk Study: The authors are grateful to the study families their participation, and to members of the E-Risk team for their dedication, hard work, and insights. The E-Risk Study is funded by the Medical Research Council (G1002190) and the National Institute of Child Health and Human Development
(HD077482). Additional support was provided by a Distinguished Investigator Award from the American Asthma Foundation to Dr. Mill, the Jacobs Foundation and the Avielle Foundation. Dr. Arseneault is the Mental Health Leadership Fellow for the UK ESRC. Dr. Agnew-Blais is an MRC Skills Development Fellow. This work used a high-performance computing facility partially supported by grant 2016-IDG-1013 ("HARDAC+: Reproducible HPC for Next-generation Genomics") from the North Carolina Biotechnology Center.
BIOS Consortium (Biobank-based Integrative Omics Study)

**Management Team** Bastiaan T. Heijmans (chair), Peter A.C. ‘t Hoen, Joyce van Meurs, Aaron Isaacs, Rick Jansen, Lude Franke.

**Cohort collection** Dorret I. Boomsma, René Pool, Jenny van Dongen, Jouke J. Hottenga (Netherlands Twin Register); Marleen MJ van Greevenbroek, Coen D.A. Stehouwer, Carla J.H. van der Kallen, Casper G. Schalkwijk (Cohort study on Diabetes and Atherosclerosis Maastricht); Cisca Wijmenga, Lude Franke, Sasha Zhernakova, Ettje F. Tigchelaar (LifeLines Deep); P. Eline Slagboom, Marijn Verkerk, Ruud van der Breggen, Jeroen van Rooij.

**Data Generation** Joyce van Meurs (Chair), P. Mila Jhamai, Michael Verbiest, H. Eka D. Suchiman, Marijn Verkerk, Ruud van der Breggen, Jeroen van Rooij.

**Data management and computational infrastructure** Hailiang Mei (Chair), Maarten van Iterson, Michiel van Galen, Jan Bot, Dasha V. Zhernakova, Rick Jansen, Peter van ‘t Hof, Patrick Deelen, Irene Nooren, Peter A.C. ‘t Hoen, Bastiaan T. Heijmans, Matthijs Moed.

**Data Analysis Group** Lude Franke (Co-Chair), Martijn Vermaat, Dasha V. Zhernakova, René Luijk, Marc Jan Bonder, Maarten van Iterson, Patrick Deelen, Freerk van Dijk, Michiel van Galen, Wibowo Arindrarto, Szymon M. Kielbasa, Morris A. Swertz, Erik. W van Zewt, Rick Jansen, Peter-Bram ‘t Hoen (Co-Chair), Bastiaan T. Heijmans (Co-Chair).

1. Molecular Epidemiology Section, Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, The Netherlands
2. Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands
3. Department of Internal Medicine, ErasmusMC, Rotterdam, The Netherlands
4. Department of Genetic Epidemiology, ErasmusMC, Rotterdam, The Netherlands
5. Department of Psychiatry, VU University Medical Center, Neuroscience Campus Amsterdam, Amsterdam, The Netherlands
6. Department of Genetics, University of Groningen, University Medical Centre Groningen, Groningen, The Netherlands
7. Department of Biological Psychology, VU University Amsterdam, Neuroscience Campus Amsterdam, Amsterdam, The Netherlands
8. Department of Internal Medicine and School for Cardiovascular Diseases (CARIM), Maastricht University Medical Center, Maastricht, The Netherlands
9. Department of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, The Netherlands
10. Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands
11. Department of Epidemiology, ErasmusMC, Rotterdam, The Netherlands
12. Sequence Analysis Support Core, Leiden University Medical Center, Leiden, The Netherlands
13. SURFsara, Amsterdam, the Netherlands
14. Gonomics Coordination Center, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands
15. Medical Statistics Section, Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, The Netherlands
Disclosures

Dr. Franke has received educational speaking fees from Shire and Medice. All other authors report no biomedical financial interests or potential conflicts of interest.

Data availability

The HumanMethylation450 BeadChip and RNA-seq data from the BIOS consortium are available in the European Genome-phenome Archive (EGA), under the accession code EGAD00010000887. Data from the Dunedin Multidisciplinary Health and Development are available via a managed access system (contact: ac115@duke.edu). The HumanMethylation450 BeadChip DNA methylation data from the E-Risk Study are available in GEO under accession number GSE105018.
References


methylolation signatures of educational attainment. npj Sci Learn. 3: 7.


Tables

Table 1 Cohort characteristics

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>NTR (N=2258*)</th>
<th>Dunedin Study (N=800)</th>
<th>E-risk study (N=1631)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAARS ADHD index</td>
<td>7.9 (3.9)</td>
<td>DSM-5 Adult ADHD Symptoms 1.5 (2.3)</td>
<td>DSM-5 Adult ADHD Symptoms 5.8 (4.3)</td>
</tr>
<tr>
<td>Age at blood sampling (years)</td>
<td>37.3 (12.9)</td>
<td>Age at blood sampling (years) 38 (0)</td>
<td>Age at blood sampling (years) 18 (0)</td>
</tr>
<tr>
<td>Sex: female</td>
<td>1549 (69.0%)</td>
<td>Sex: female 392 (49.0%)</td>
<td>Sex: female 817 (50.1%)</td>
</tr>
<tr>
<td>Current smoker</td>
<td>438 (19.4%)</td>
<td>Current smoker 162 (20.3%)</td>
<td>Current smoker** 372 (22.8%)</td>
</tr>
<tr>
<td>Former smoker</td>
<td>532 (23.6%)</td>
<td>Former smoker 233 (29.1%)</td>
<td></td>
</tr>
<tr>
<td>Never smoked</td>
<td>1288 (57.0%)</td>
<td>Never smoked 405 (50.6%)</td>
<td>Non-smoker ** 1259 (77.2%)</td>
</tr>
</tbody>
</table>

* In NTR, 2258 samples from 2232 individuals were included (for 26 individuals, two longitudinal DNA samples were included).

** In E-risk, the smoking variable that was included as covariate in the analyses and that is presented in this table is current daily smoking at age 18.
Table 2 Top DMPs from the EWAS meta-analysis.

<table>
<thead>
<tr>
<th>cgID</th>
<th>CHR</th>
<th>Position*</th>
<th>Gene</th>
<th>Location</th>
<th>Nearest Gene</th>
<th>Weight</th>
<th>Z score</th>
<th>P</th>
<th>Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>cg26197679</td>
<td>8</td>
<td>142310085</td>
<td>Intergenic</td>
<td>LINC01300</td>
<td>4689</td>
<td>-4.9</td>
<td>1.1 x 10^-6</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>cg23144852</td>
<td>3</td>
<td>193405999</td>
<td>OPA1</td>
<td>Gene body</td>
<td>OPA1</td>
<td>4684</td>
<td>-4.8</td>
<td>1.7 x 10^-6</td>
<td>---</td>
</tr>
<tr>
<td>cg10984962</td>
<td>2</td>
<td>236462202</td>
<td>AGAP1</td>
<td>Gene body</td>
<td>AGAP1</td>
<td>4688</td>
<td>4.5</td>
<td>9.7 x 10^-6</td>
<td>+++</td>
</tr>
</tbody>
</table>

note: CpGs with a p-value lower than 1.0 x 10^-5 are shown. CHR=chromosome. P=p-value. Weight=Total sample size in the meta-analysis. Z score=Meta-analysis Z score. *Genome build Hg19 (build 37).
Table 3 Significant DMRs associated with ADHD symptoms

<table>
<thead>
<tr>
<th>chr</th>
<th>start</th>
<th>end</th>
<th>N probes</th>
<th>region size (bp)</th>
<th>Region Šidák p-value</th>
<th>gene (s) associated with CpGs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Netherlands Twin Register (NTR)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>31650735</td>
<td>31651363</td>
<td>19</td>
<td>628</td>
<td>4.3x10^{-10}</td>
<td>LY6G5C, DDAH2, AIF1, C6orf48, LST1, VWA7, HSPA1L, PRRC2A, MICB</td>
</tr>
<tr>
<td>11</td>
<td>67417958</td>
<td>67418406</td>
<td>11</td>
<td>448</td>
<td>7.6x10^{-9}</td>
<td>ACY3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>31583915</td>
<td>31584224</td>
<td>4</td>
<td>309</td>
<td>1.2x10^{-6}</td>
<td>AIF1</td>
</tr>
<tr>
<td>6</td>
<td>33245460</td>
<td>33246308</td>
<td>32</td>
<td>848</td>
<td>3.0x10^{-8}</td>
<td>B3GALT4, B3GALT4, RPS18</td>
</tr>
<tr>
<td>13</td>
<td>113613574</td>
<td>113613738</td>
<td>3</td>
<td>164</td>
<td>4.8x10^{-2}</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>48632484</td>
<td>48633145</td>
<td>11</td>
<td>661</td>
<td>2.1x10^{-4}</td>
<td>COL7A1, COL7A1</td>
</tr>
<tr>
<td><strong>Dunedin Study</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>32063394</td>
<td>32065212</td>
<td>51</td>
<td>1818</td>
<td>2.3 x 10^{-17}</td>
<td>SKIV2L, C4B, C4A, TNXB, TNXA</td>
</tr>
<tr>
<td>5</td>
<td>191127</td>
<td>192104</td>
<td>12</td>
<td>977</td>
<td>4.4 x 10^{-11}</td>
<td>LRRC14B</td>
</tr>
<tr>
<td>19</td>
<td>17931932</td>
<td>17932370</td>
<td>8</td>
<td>438</td>
<td>2.3 x 10^{-8}</td>
<td>INSL3</td>
</tr>
<tr>
<td>20</td>
<td>748992</td>
<td>749621</td>
<td>8</td>
<td>629</td>
<td>4.9 x 10^{-6}</td>
<td>C20orf54</td>
</tr>
<tr>
<td>6</td>
<td>32847441</td>
<td>32847846</td>
<td>16</td>
<td>405</td>
<td>1.3 x 10^{-6}</td>
<td>PPP1R2P1, HLA-DQA</td>
</tr>
<tr>
<td>15</td>
<td>75019070</td>
<td>75019377</td>
<td>10</td>
<td>307</td>
<td>4.8 x 10^{-6}</td>
<td>CYP1A1</td>
</tr>
<tr>
<td>4</td>
<td>56458183</td>
<td>56458623</td>
<td>7</td>
<td>440</td>
<td>5.1 x 10^{-6}</td>
<td>PDCL2</td>
</tr>
<tr>
<td>11</td>
<td>1769152</td>
<td>1769646</td>
<td>9</td>
<td>494</td>
<td>6.8 x 10^{-6}</td>
<td>HCCA2; LOC40277B; LOC40277B</td>
</tr>
<tr>
<td>4</td>
<td>186732837</td>
<td>186733332</td>
<td>8</td>
<td>495</td>
<td>8.9 x 10^{-6}</td>
<td>SORBS2</td>
</tr>
<tr>
<td>15</td>
<td>90727560</td>
<td>90728141</td>
<td>9</td>
<td>581</td>
<td>5.5 x 10^{-5}</td>
<td>SEMA4B, SEMA4B</td>
</tr>
<tr>
<td>7</td>
<td>1250038</td>
<td>1250757</td>
<td>10</td>
<td>719</td>
<td>6.4 x 10^{-5}</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2728344</td>
<td>2728913</td>
<td>4</td>
<td>569</td>
<td>8.8 x 10^{-5}</td>
<td>AMZ1</td>
</tr>
<tr>
<td>8</td>
<td>41655673</td>
<td>41656080</td>
<td>6</td>
<td>407</td>
<td>1.5 x 10^{-4}</td>
<td>ANK1</td>
</tr>
<tr>
<td>12</td>
<td>25801455</td>
<td>25801622</td>
<td>4</td>
<td>167</td>
<td>3.5 x 10^{-4}</td>
<td>IFLTD1</td>
</tr>
<tr>
<td>6</td>
<td>3247958</td>
<td>3248099</td>
<td>2</td>
<td>141</td>
<td>2.3 x 10^{-3}</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>96670628</td>
<td>96671170</td>
<td>11</td>
<td>542</td>
<td>3.9 x 10^{-3}</td>
<td>BDKRB2</td>
</tr>
<tr>
<td>20</td>
<td>22795685</td>
<td>22795775</td>
<td>2</td>
<td>90</td>
<td>8.7 x 10^{-3}</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>187985377</td>
<td>187985473</td>
<td>2</td>
<td>96</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1993528</td>
<td>1993546</td>
<td>2</td>
<td>18</td>
<td>0.03</td>
<td>MYOM2</td>
</tr>
</tbody>
</table>
chr= chromosome, N probes= total number of methylation sites in the window, bp= base pairs, gene(s)= gene(s) mapping to CpGs in the window. More detailed information on the regions is provided in Table S7. More detailed results for the association between DNA methylation levels and gene expression in cis is provided in Table S8 (NTR) and Table S9 (Dunedin Longitudinal Study). No significant DMRs were found in E-risk.
### Primary analyses

<table>
<thead>
<tr>
<th>Study-specific EWAS analyses</th>
<th>NTR N=2258</th>
<th>Dunedin N=800</th>
<th>E-risk N=1631</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meta-analysis of 394 194 CpGs N=4689</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Secondary analyses

<table>
<thead>
<tr>
<th>Enrichment</th>
<th>Regions</th>
<th>Follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>• GWAS loci • EWAS loci</td>
<td>• Study-specific DMR analysis • Comparison of DMRs across cohorts • Meta-analysis-based DMR analysis</td>
<td>• Gene expression • mQTLs • Blood-brain correlation</td>
</tr>
</tbody>
</table>
Epigenome-wide Association Study of Attention-Deficit/Hyperactivity Disorder Symptoms in Adults

Supplement 1

EWAS Model Equations

The following models were fitted in each cohort:

**NTR**

\[ CpGi = \alpha + \beta_{ADHDsymptoms} \times ADHDsymptoms + \beta_{sex} \times sex + \beta_{age} \times age + \beta_{smoking} \times smoking + \beta_{gPC1} \times gPC1 + \beta_{gPC2} \times gPC2 + \beta_{gPC3} \times gPC3 + \beta_{monocyte} \times monocyte\% + \beta_{eosinophil} \times eosinophil\% + \beta_{neutrophil} \times neutrophil\% + \beta_{arrayrow} \times arrayrow + \beta_{sampleplate} \times plate \ldots + \beta_{sampleplate33} \times plate33 + \epsilon \]

**Dunedin**

\[ CpGi = \alpha + \beta_{ADHDsymptoms} \times ADHDsymptoms + \beta_{sex} \times sex + \beta_{smoking} \times smoking + \beta_{gPC1} \times gPC1 + \beta_{gPC2} \times gPC2 + \beta_{monocyte} \times monocyte\% + \beta_{eosinophil} \times eosinophil\% + \beta_{neutrophil} \times neutrophil\% + \beta_{mPC1} \times mPC1 \ldots + \beta_{mPC32} \times mPC32 + \epsilon \]

**E-Risk**

\[ CpGi = \alpha + \beta_{ADHDsymptoms} \times ADHDsymptoms + \beta_{sex} \times sex + \beta_{smoking} \times smoking + \beta_{gPC1} \times gPC1 + \beta_{gPC2} \times gPC2 + \beta_{plasmablast} \times plasmablast\% + \beta_{CD8+CD28-CD45RA-} \times CD8 + CD28 - CD45RA - T\% + \beta_{CD8T} \times CD8T\% + \beta_{CD4T} \times CD4T\% + \beta_{NK} \times NK\% + \beta_{monocyte} \times monocyte\% + \beta_{granulocyte} \times granulocyte\% + \beta_{mPC1} \times mPC1 \ldots + \beta_{mPC28} \times mPC28 + \epsilon \]

where \( CpGi \) is DNA methylation \( \beta \)-value at methylation site \( i \), \( \alpha \) is the intercept, \( ADHDsymptoms \) is ADHD symptoms, sex is coded as 0 for males and 1 for females, age is the age at blood sampling in years (which only varies across individuals in NTR), smoking is smoking status, coded as 0 (never smoked), 1 (former-smoker), 2 (current-smoker) in NTR and Dunedin and coded as 0 (never-smoked) and 1 (current-smoker) in E-Risk, gPC1, gPC2, and gPC3 are PCs 1, 2, and 3 based on genome-wide genotype data, respectively,
monocyte%, eosinophil% and neutrophil%, plasmablast%, CD8+CD28-CD45RA- T%, CD8T%, CD4T%, NK%, and granulocytes%, are percentages of white blood cells, array row is the row of the sample on the Illumina 450k Beadchip (ranging from 1 to 6), plate1...plate33 are 96-wells bisulfite plates (the samples were processed on 34 plates in total), mPC1 ... mPC32, are Principal Components 1 to 32 based on methylation array control probes, and \( \varepsilon \) is residual.

Cohort 1: Netherlands Twin Register

Subjects and samples

The subjects participated in longitudinal survey studies from the Netherlands Twin Register (NTR) (1) and were phenotyped for ADHD symptoms multiple times. Between 2004 and 2011, a subsample was invited to take part in the NTR biobank project (2). The longitudinal survey design and blood sampling procedures have been described in detail previously (1, 2). In the current EWAS, we included individuals for whom the following data were available: phenotype information from the CAARS ADHD index, good quality DNA methylation data, genome-wide SNP data, white blood cell percentages (monocytes, neutrophils and eosinophils), and smoking status, and excluded 'ethnic outliers'; individuals who were outliers on Principal Components (PCs) based on genome-wide genotype data. This left 2258 samples from 2232 subjects (for 26 subjects, data from two longitudinal methylation samples were included). Informed consent was obtained from all participants. The study was approved by the Central Ethics Committee on Research Involving Human Subjects of the VU University Medical Centre, Amsterdam, an Institutional Review Board certified by the U.S. Office of Human Research Protections (IRB number IRB00002991 under Federal-wide Assurance FWA00017598; IRB/institute codes, NTR 03-180).

ADHD symptoms

Data on ADHD symptoms were collected in multiple NTR surveys. For the EWAS, data from the CAARS (Conners' Adult ADHD Rating Scales) ADHD index (3), collected in NTR Survey 7 (data collection in 2004; \( N=1225 \)), Survey 8 (data collection in 2009; \( N=875 \)), and Survey 10 (data collection in 2013; \( N=132 \)) were analyzed. The CAARS ADHD index assesses total ADHD symptoms based on 12 items, each of which is scored on a scale from 0 to 4. Thus, the possible range of the CAARS is 0 to 48. For each participant, the ADHD index obtained closest to the date of blood draw was selected. The average time between blood sampling and ADHD symptom assessment was 1.9 years (median=1.5, range: -9.6 years (blood sample
before survey) ~ 13.1 years (blood sample after survey)). **Supplemental Figure S1a** shows the distribution of the ADHD scores in the 2232 subjects included in the analysis. The CAARS ADHD index is designed to identify adults who are likely to be diagnosed with ADHD. If ADHD index scores are transformed into sex-specific T-scores, a T-score >65 is considered significant in clinical groups (3). The prevalence of ADHD in the NTR based on the T-score cut-off of >65 has been previously described and was 6.8% in women and 7.4% in men (4). Sensitivity analyses were performed on data from the full 30-item CAARS version, included in NTR survey 7, available for 1846 samples. This version includes, in addition to total ADHD symptoms, a 9-item subscale for inattentive symptoms (inattention), and a 9-item subscale for hyperactive-impulsive symptoms (hyperactivity), corresponding to the symptoms that represent the diagnostic criteria of adult ADHD as outlined in DSM-IV-TR. Items from the three subscales did not overlap.

**Peripheral blood DNA methylation**

DNA methylation was assessed with the Infinium HumanMethylation450 BeadChip Kit (HM450k; Illumina, San Diego, CA, USA) by the Human Genotyping facility (HugeF) of ErasmusMC, The Netherlands [http://www.glimdna.org/](http://www.glimdna.org). Genomic DNA (500 ng) from whole blood was bisulfite-treated using the Zymo EZ DNA Methylation kit (Zymo Research Corp, Irvine, CA, USA), and 4 µl of bisulfite-converted DNA was measured on the Illumina 450k array following the manufacturer’s protocol. A number of sample- and probe-level quality checks and sample identity checks were performed (described in detail previously (5)). In short, sample-level QC was performed using MethylAid (6). Probes were set to missing in a sample, if they had an intensity value of exactly zero, a detection \( p > .01 \), or a bead count of <3. After these steps, probes that failed based on the above criteria in >5% of the samples were excluded from all samples (only probes with a success rate ≥0.95 were retained). The methylation data were normalized with functional normalization (7), and normalized intensity values were converted into beta (\( \beta \))-values.

**Covariates**

White blood cell percentages were included as covariates in the EWAS to account for variation in cellular composition between whole blood samples, and were obtained as part of the complete blood count (2). The following WBC were included as covariates: monocytes, eosinophils, and neutrophils (lymphocyte percentage was not included because it correlated with neutrophils (\( r = -0.9 \)), and basophil percentage was not included because it showed very little variation between individuals). Information on current and past smoking behavior was collected as part of the NTR biobank project at the moment of blood draw. Smoking status was coded as 0 (never smoked), 1 (former smoker), 2 (current smoker). HM450k array row and
bisulfite plate (dummy-coding) were included as covariates to account for technical variation. The first three PCs based on genome-wide SNP data, which reflect population structure within the Netherlands (8), were included as covariates to account for population differences. Finally, age and sex were included as covariates in NTR, to account for variation in age and sex.

**Epigenome-wide association study (EWAS)**

EWAS analyses were performed in R software. The association between DNA methylation level and CAARS index was tested with DNA methylation \( \beta \)-value as outcome and predictors CAARS index score, sex, age at blood sampling, smoking status, percentages of monocytes, eosinophils, and neutrophils, HM450k array row, bisulfite plate, and PC1, PC2 and PC3 from genome-wide genotype data. EWAS analyses were performed with generalized estimation equation (GEE) models, which were fitted with the R package ‘gee’. The following settings were used: Gaussian link function (for continuous data), 100 iterations, and the ‘exchangeable’ option to account for the correlation structure within families and within persons.

**Cohort 2: Dunedin Multidisciplinary Health and Development Study**

**Subjects and samples**

Participants are members of the Dunedin Multidisciplinary Health and Development Study, a longitudinal investigation of the health and behavior of a representative birth cohort of consecutive births between April 1972 and March 1973 in Dunedin, New Zealand. The cohort of 1,037 children (52% boys) was constituted at age 3 as 91% of eligible births resident in the province. The cohort represents the full range of socioeconomic status on NZ’s South Island and matches the NZ National Health and Nutrition Survey on adult health indicators (e.g., BMI, smoking, GP visits) (9). Cohort members are primarily white; approximately 7% self-identify as having any non-white ancestry, matching the South Island. Follow-up assessments were conducted at ages 5, 7, 9, 11, 13, 15, 18, 21, 26, 32, and most recently 38 years, when 95% of the 1,007 living study members underwent assessment in 2010-2012. DNA methylation data were generated for a subset of this cohort (non-Maori participants). In the current EWAS, we included individuals for whom the following data were available: ADHD symptom scores, good quality DNA methylation data, white blood cell counts, smoking status, and genome-wide SNP data, leaving 800 subjects in the analysis sample. Information on current and past smoking behavior was collected on the
same day of blood draw. The study protocol was approved by the institutional ethical review boards of the participating universities. Study members gave informed consent before participating.

**ADHD symptoms**

ADHD symptoms were assessed at age 38 years, at the same time point as blood collection. Ascertainment of ADHD Symptoms is described in Moffitt et al. (2015) (10). Briefly, symptoms were ascertained in 2010-2012 through private structured diagnostic interviews by interviewers with mental-health-related tertiary qualifications and clinical experience. Interviewers received 2 weeks of formal training on the mental health interview, and were trained to inter-rater reliability criterion standard for ascertainment of symptoms, plus re-training continued periodically to prevent drift. Age-38 interviewers were blind to prior data. Because the Dunedin Study's age-38 assessments began in 2010, question/items administered were those reported in leading measures of adult ADHD considered by the DSM-5 working group at that time (11–13). Our interview format followed recommendations from the working group to include behavioral examples relevant for adults (for example, if an item referred to jumping out of seat for children, the interviewer gave examples of difficulty sitting through long meetings or social occasions for adults). The reporting period was the past 12 months. Responses were scored 0=NO/1=YES, and scores summed across the 18 items. To receive an adult ADHD diagnosis, DSM-5 requires that 5 or more inattentive and/or 5 or more hyperactivity-impulsivity symptoms are present. Additional details regarding DSM-5 ADHD symptom scores, how they translate into an ADHD diagnosis, and the prevalence of ADHD (3% in the Dunedin study at age 38 years) have been described previously (10).

**Peripheral blood DNA methylation**

DNA methylation data were generated using the Infinium HumanMethylation450 BeadChips (Illumina, CA, USA). Whole-blood was collected from the non-Maori participants in K$_2$EDTA vacutainer tubes (BD, NJ, USA). DNA was extracted from the buffy coat using standard procedures (14, 15). ~500ng of DNA from each sample was treated with sodium bisulfite using the EZ-96 DNA Methylation kit (Zymo Research, CA, USA). Array analysis was performed by the Duke University Molecular Physiology Institute Genomics Core Facility using the iScan platform (Illumina).

Data were processed and normalized using the methylumi (v2.14.0) Bioconductor package from the R statistical programming environment, and subjected to quality control analyses. Briefly, the method corrects Cy3 and Cy5 dye bias and recalculates the betas based on the corrected intensities against a reference array, which defaults to the first chip in the set. Samples were removed if the average detection
p-value was >= 0.001. To confirm genetic identity of the DNA samples, we assessed genotype concordance between SNP probes on the 450K array and data generated using Illumina OmniExpress12v1.1 genotyping BeadChips.

**Covariates**

Percentages of neutrophils, monocytes, and eosinophils were included as covariates in the EWAS to account for variation in cellular composition between whole blood samples, and were measured using flow cytometry (Sysmex Corporation, Japan) in whole blood samples taken concurrently with the DNA sample. Information on smoking status was obtained at the moment of blood draw and coded as 0 (never smoked), 1 (former smoker), 2 (current smoker). To permit control for technical variation, we used methylation-array control-probe principal components (16). 32 principal components were needed to explain 90% of the variance. These principal components were used as covariates in the analyses. Age was not included as a covariate because blood sample collection and ADHD symptom assessment took place, for all subjects, at the same age (38 years).

**Epigenome-wide association study (EWAS)**

EWAS analyses were performed in R software. The association between DNA methylation level and ADHD symptoms was tested under a linear model with DNA methylation β-value as outcome and predictors ADHD symptoms, sex, smoking status, neutrophil percentage, monocyte percentage, eosinophil percentage, and PCs1-32 (from genome-wide methylation control probe data), and PCs 1 and 2 from genome-wide SNP data. EWAS models were fitted with the R function lm().

**Cohort 3: E-risk Twin Study**

**Subjects and samples**

Participants were members of E-Risk, which tracks the development of a 1994-95 birth cohort of 2,232 British children (17). Briefly, the E-Risk sample was constructed in 1999-2000, when 1,116 families (93% of those eligible) with same-sex 5-year-old twins participated in home-visit assessments. This sample comprised 56% monozygotic (MZ) and 44% dizygotic (DZ) twin pairs; sex was evenly distributed within zygosity (49% male). The study sample represents the full range of socioeconomic conditions in Great Britain, as reflected in the families’ distribution on a neighborhood-level socioeconomic index (ACORN [A
Classification of Residential Neighbourhoods], developed by CACI Inc. for commercial use): 25.6% of E-Risk families live in "wealthy achiever" neighborhoods compared to 25.3% nationwide; 5.3% vs. 11.6% live in "urban prosperity" neighborhoods; 29.6% vs. 26.9% in "comfortably off" neighborhoods; 13.4% vs. 13.9% in "moderate means" neighborhoods; and 26.1% vs. 20.7% in "hard-pressed" neighborhoods. E-Risk underrepresents "urban prosperity" neighborhoods because such households are often childless.

Home visits were conducted when participants were aged 5, 7, 10, 12 and most recently, 18 years (93% participation). At age 18, whole blood was collected from 82% (N=1700) of the participants. In the current EWAS, we included individuals for whom the following data were available: ADHD symptom scores, good quality DNA methylation data, smoking status, and genome-wide SNP data leaving 1631 individuals. The Joint South London and Maudsley and the Institute of Psychiatry Research Ethics Committee approved each phase of the study. Study members gave informed consent before participating.

**ADHD symptoms**

ADHD symptoms were assessed at age 18 years, at the same time point as blood collection. Assessment of ADHD symptoms at age 18 years is described in Agnew-Blais et al (18). Briefly, symptoms were assessed based on private structured interviews with participants regarding 18 symptoms of inattention and hyperactivity-impulsivity according to DSM-5 criteria. Responses were coded 0=NO and 1=YES; and scores were summed across the 18 items. To receive an adult ADHD diagnosis, DSM-5 requires that 5 or more inattentive and/or 5 or more hyperactivity-impulsivity symptoms are present. Additional details regarding DSM-5 ADHD symptom scores, how they translate into ADHD diagnosis, and the prevalence of ADHD (8% in E-Risk at age 18 years) have been described previously (18).

**Peripheral blood DNA methylation**

DNA methylation data were generated using the Infinium HumanMethylation450 BeadChips run on an Illumina iScan System (Illumina, CA, USA). Whole blood was collected in 10mL K$_2$EDTA tubes and DNA extracted from the buffy coat using a Flexigene DNA extraction kit (Qiagen, Hilden, Germany) following manufacturer's instructions.

We assayed 1669 blood samples (out of 1700); 31 samples were not useable (e.g., due to low DNA concentration). ~500ng of DNA from each sample was treated with sodium bisulfite using the EZ-96 DNA Methylation kit (Zymo Research, CA, USA). Twin pairs were randomly assigned to bisulfite-conversion plates and Illumina 450K arrays, with siblings processed in adjacent positions to minimize batch effects.
Fully methylated control samples (CpG Methylated HeLa Genomic DNA; New England BioLabs, MA, USA) were included in a random position on each plate; the distinct DNA methylation profile of this sample enabled us to confirm the experiment was successful and to ensure there were no plate mix-ups or rotations.

Data were imported using the `methylumIDAT` function in `methylum` (19), and subjected to quality control analyses. First, we excluded all samples with median methylated (‘M’) and unmethylated (‘U’) intensities <2500. Second, using the ten control probes included on the 450K array, we examined the efficiency of the sodium bisulfite conversion reaction; samples were excluded if their "conversion score" was <80. Third, multidimensional scaling was performed for DNA methylation probes on each of the sex chromosomes and compared to the reported gender. Fourth, to confirm genetic identity of the DNA samples, we assessed genotype concordance between SNP probes on the 450K array and data generated using Illumina OmniExpress24v1.1 genotyping BeadChips.

Samples from 1658 participants passed our QC pipeline. Data were processed with the `pfilter` function from the `watermelon` package (Pidsley et al., 2013) excluding 0 samples with >1% of sites with a detection p value >0.05, 567 sites with beadcount <3 in 5% of samples and 1448 probes with >1% of samples with detection p value >0.05. The data were normalized with the `dasen` function from the `watermelon` package (20).

**Covariates**

To control for cell type composition, we used as covariates cell-type proportions estimated from the methylation data (21). Information on smoking status was obtained on the day of blood draw and coded as 0 (never smoked), 1 (current smoker). To permit control for technical variation, we used methylation-array control-probe principal components (16). 28 principal components were needed to explain 90% of the variance. These principal components were used as covariates in the analyses. Age was not included as a covariate because blood sample collection and ADHD symptom assessment took place, for all subjects, at the same age (18 years).

**Epigenome-wide association study (EWAS)**

EWAS analyses were performed in R software. The association between DNA methylation level and ADHD symptoms was tested with DNA methylation β-value as outcome and predictors ADHD symptoms, sex,
current smoking status, PCs 1-28 (from genome-wide methylation control probe data), estimated cell counts (plasma blasts, CD8+CD28-CD45RA- T cells, naïve CD8 T cells, CD4 T cells, natural killer cells, monocytes, granulocytes), and PCs 1 and 2 from genome-wide SNP data. EWAS analyses were performed with generalized estimation equation (gee) models, which were fitted with the R package ‘gee’. The following settings were used: Gaussian link function (for continuous data), 100 iterations, and the ‘exchangeable’ option to account for the correlation structure within twin pairs.

Motivation for Covariates in the EWAS Analysis

Smoking has widespread effects on DNA methylation (22), and individuals with more ADHD symptoms smoke more (23). We adjusted for smoking status to control for confounding effects of smoking, although we recognize the disadvantage of potentially removing relevant ADHD-related variation. WBC counts were measured in NTR and Dunedin; the following were included as covariates: monocytes, eosinophils, and neutrophils. In E-risk, WBCs were predicted because measured counts were not available; the following were included as covariates: plasma blasts, CD8+CD28-CD45RA- T cells, naïve CD8 T cells, CD4 T cells, natural killer cells, monocytes, granulocytes. Since the molecular contributions to ADHD may vary across individuals with different genetic backgrounds (24), we excluded individuals for whom Principal Components (PCs) based on genome-wide SNPs indicated distinct ancestry (only applicable to NTR), and we included PCs based on genome-wide SNP data as covariates in the EWAS model to account for population structure within each population.

Enrichment Analyses

To study overlap of EWAS signal from the meta-analysis with genetic findings for ADHD, we considered the most recent GWAS for ADHD (25). Since the sample size of the ADHD GWAS is relatively small, and because there appears to be considerable overlap of genetic effects across psychiatric disorders (26), we also considered findings from the much larger GWASs of major depressive disorder (MDD) (27) and schizophrenia (28), and from the GWAS of autism spectrum disorders (ASD) (29). ADHD GWAS summary statistics were obtained from the PGC/iPSYCH consortium. MDD2 GWAS summary statistics (based on the meta-analysis without NTR) were obtained from the PGC. Schizophrenia GWAS summary statistics were obtained from the PGC website (28). ADD summary statistics from the PGC/iPSYCH consortium were
obtained from the iPSYCH website. For each GWAS, we obtained a pruned SNP list, from which we selected all SNPs with a p-value < 1.0x10^-4 and determined the distance of each Illumina 450k methylation site to each SNP. We considered three windows around GWAS loci, 10 kb, 100 kb, and 1 Mb, to annotate methylation sites, and applied two p-value thresholds to select GWAS loci (genome-wide significant SNPs; p<5.0x10^-8 and nominally significant SNPs; p<1.0x10^-4). The top 100 CpGs associated with ADHD symptom trajectories in childhood were obtained from Walton et al. (30). Because of the small number of epigenome-wide significant loci for the ADHD EWAS, we analysed the top 100 CpGs. Methylation sites associated with schizophrenia were obtained from Hannon et al. (31), selecting all epigenome-wide significant CpGs (p<1.0x10^-7) from their meta-analysis of all cohorts. Methylation sites previously associated with individual smoking or prenatal exposure to maternal smoking at a false-discovery rate (FDR) < 5.0% were obtained from recent large meta-analyses by Joehanes et al. (22) and Joubert et al. (32). To test whether a CpG category was enriched among sites more strongly associated with ADHD symptoms, meta-analysis EWAS test statistics were regressed on each CpG category:

\[ \text{\mid Z_{score} \mid = \text{Intercept} + \beta_{category \_x} \times \text{Category } x} \]

Where \( \text{\mid Z_{score} \mid} \) represents the absolute Zscore from the EWAS meta-analysis of ADHD symptoms; \( \beta_{category \_x} \) represents the estimate for category x, i.e., the change in the EWAS test statistic associated with a one unit change in category x (e.g. being located within 100 kb of SNPs associated with ADHD). For each enrichment test, bootstrap standard errors were computed with 2000 bootstraps with the R-package ‘simpleboot’.

Statistical significance was assessed after Bonferroni correction for the total number of GWASs of psychiatric traits (N=4) and EWASs of traits and exposures (N=4); alpha=0.05/8=0.006.

**Differentially Methylated Regions**

We used the python module 'Comb-p'(33) to scan for regions in which multiple correlated methylation sites showed evidence for association with ADHD symptoms. Comb-p corrects EWAS p-values for the correlation between sites within a particular window and calculates an overall p-value for the region. To consider regions of variable sizes, we applied a step-wise approach, starting with correlated CpGs within an arbitrarily chosen window of 50 bp (seed P-value < 0.01, minimum of 2 probes) and then stepwise increasing the window size up till 1500 (or until no more significant regions were detected). In this, we
considered the window sizes 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, and 1500 bp. Šidák correction, as implemented in Comb-p, was applied to calculate the p-values for DMRs. For each DMR, Šidák correction accounts for the number of possible tests, defined as the total bases covered by all input probes divided by the size of the region. We report significant regions (Šidák p<0.05) with at least two methylation sites within a 500 bp window. Comb-p was applied to the EWAS results from each of the three cohorts, separately, and on the EWAS meta-analysis. DMR analysis may detect loci where multiple CpGs are associated with ADHD symptoms, but where the individual associations are not significant in EWAS due to insufficient power. Regional plots of DMRs were created with coMET (34).

**Gene Expression in Cis**

We examined if DNA methylation is associated with gene expression levels in cis, using an independent whole blood RNA-sequencing dataset (which did not include NTR participants) from the Biobank-based Integrative Omics Study (BIOS) consortium that was described previously (35). This study tested associations between genome-wide CpGs and transcripts in cis (<250 kb). In short, methylation and expression levels in whole-blood samples (n=2,101) were quantified with Illumina Infinium HumanMethylation450 BeadChip arrays and with RNA-seq (2x50bp paired-end, Hiseq2000, >15M read pairs per sample). For each target CpG (sites with ADHD symptom EWAS meta-analysis p<1x10^{-5} or located in significant DMRs), we identified transcripts in cis (<250 kb), for which methylation levels were significantly associated with gene expression levels at the experiment-wide threshold applied by this study (FDR<5.0%), after regressing out mQTL and eQTL effects. None of the samples included in the EWAS of ADHD symptoms was included in the expression analysis.

**mQTLs**

Previously published data from the BIOS consortium were used to look up if top-DMPs and CpGs in DMRs were significantly associated with mQTLs in blood, at the experiment-wide threshold applied by this study (FDR<5.0%, (35)). This study tested both cis and trans mQTL relationships and was performed on 3841 peripheral blood samples (including a subset of the samples from NTR).
Correlation Between DNA Methylation Level in Blood and Brain

Previously published correlations between DNA methylation levels in blood and four brain regions from matched samples (prefrontal cortex, entorhinal cortex, superior temporal gyrus and cerebellum) were obtained from Hannon et al (36). Statistical significance was assessed after Bonferroni correction for the number of brain regions (=4) multiplied by the number of DMPs (=3) plus the number of DMRs tested (=25): 4 * 28=112, giving an alpha of 0.05/112=4.46x10^-4. The online tool was used to plot DNA methylation levels in blood and brain (https://epigenetics.essex.ac.uk/bloodbrain/).

Power Analysis

Power analysis was performed to estimate the sample size required to achieve 80% power to detect the associations between methylation and ADHD symptoms for the three top-DMPs from the meta-analysis at an alpha of 1.0x10^-7 (epigenome-wide significance), given their effect size observed in NTR. Power analysis was performed with the function pwr.f2.test() from the R-package pwr. The percentage of variance explained ($r^2$) at each of the three ADHD top sites was obtained by squaring the correlation for CpGi ($r_i$), which was derived as follows:

$$r_i = \frac{\beta_i}{\sqrt{sdy_i^2}}$$

where $\beta_i$ is the beta from the regression of methylation level on ADHD symptoms in NTR for CpGi, $sdy_i$ is the standard deviation of DNA methylation level residuals of CpGi (obtained in NTR after adjusting methylation levels for covariates), and $sdx$ is the standard deviation of ADHD symptoms in NTR.
Supplemental Figure S1. Histograms of ADHD symptoms: a) Netherlands Twin Register, b) Dunedin Study, c) E-risk Study
Supplemental Figure S2. QQ plots.
Supplemental Figure S3. Regional plot of a DMR on chromosome 11 that is significantly associated with ADHD symptoms in NTR and with ACY3 RNA levels in the BIOS consortium (higher methylation levels correlate with lower transcript levels and with lower ADHD symptoms). The top panel shows the EWAS results for ADHD symptoms in NTR. P-values for the individual CpG sites are plotted, with the most strongly associated CpG highlighted in blue. Overall, the region is significantly associated with ADHD symptoms (p=7.6x10^-9). The middle panel shows the ACY3 gene, of which RNA levels are significantly associated with the methylation level of CpG sites in this DMR, and several regulatory tracks. The Regulatory elements track from ENSEMBL shows a promoter-associated (blue) element that overlaps with this DMR. The bottom panel shows the correlations between methylation levels of CpGs in the DMR in NTR, illustrating strong positive correlations between CpGs. The figure was created with coMET (34).
Supplemental Figure S4. Network of mQTLs and methylation sites for a DMR on chromosome 11 that is significantly associated with ADHD symptoms in NTR and with ACY3 RNA levels in the BIOS consortium. Trans-mQTL SNPs include rs11966072, rs9386791, rs9374080, and rs1008084 (chromosome 6), rs11190133 (chromosome 10), and rs2900333 (chromosome 12). All other SNPs are located in cis. The network results (SNP-CpG and CpG-RNA relationships) are based on BIOS consortium data.
Supplemental Figure S5. Example of a CpG site located in an ADHD-associated DMR on chromosome 6 with correlated DNA methylation levels in blood and brain. PFC=prefrontal cortex, EC=entorhinal cortex, STG=superior temporal gyrus, CER=cerebellum.
Supplemental Figure S6. Power analysis results for top-DMPs. The figure shows the required sample size (y-axis) to detect CpGs at an alpha of $1.0 \times 10^{-7}$, based on their effect size ($r^2$: proportion of variance explained) in NTR (x-axis). The power analysis was performed for the three top sites from the meta-analysis. Note that the true effect size and power will be different in the case of heterogeneity across cohorts.
Supplemental References


