# DNA methylation signatures of aggression and closely related constructs

## A meta-analysis of epigenome-wide studies across the lifespan

Jenny van Dongen, PhD[1]; Fiona A. Hagenbeek, Msc[1]; Matthew Suderman, PhD[2,3]; Peter Roetman, Msc[4]; Karen Sugden, PhD[5,6]; Andreas G. Chiocchetti, PhD[7]; Khadeeja Ismail, Msc[8]; Rosa H. Mulder, Msc[9,10,11]; Jonathan Hafferty, MRCPsych, PhD[12]; Mark J. Adams, PhD[12]; Rosie M. Walker, PhD[13]; Stewart W. Morris,BSc[13]; Jari Lahti, PhD[14,15]; Leanne K. Küpers, PhD[16]; Georgia Escaramis, PhD[17,18,19]; Silvia Alemany, PhD[20,21,22]; Marc Jan Bonder, PhD[23]; Mandy Meijer, Msc[24]; Hill F. Ip, Msc[1]; Rick Jansen, PhD[25]; Bart M. L. Baselmans, PhD[1]; Priyanka Parmar, MPH[26,27]; Estelle Lowry, PhD[26,28]; Fabian Streit, PhD[29]; Lea Sirignano, Msc[29]; Tabea Send, Msc[30]; Josef Frank, PhD[29]; Juulia Jylhävä, PhD[31]; Yunzhang Wang, Msc[31]; Pashupati Prasad Mishra, Msc[32]; Olivier F. Colins, PhD[4,33]; David Corcoran, PhD [6]; Richie Poulton, PhD[34]; Jonathan Mill, PhD[35]; Eilis J. Hannon, PhD[35]; Louise Arseneault, PhD[36]; Tellervo Korhonen, PhD[8]; Eero Vuoksimaa, PhD[8]; Janine Felix, PhD, MD[11,37]; Marian Bakermans-Kranenburg, PhD[38]; Archie Campbell, MA[13]; Darina Czamara, PhD[14]; Elisabeth Binder, PhD[14]; Eva Corpeleijn, PhD[16]; Juan Ramon González, PhD[20,21,22]; Regina Grazuleviciene, PhD, MD[39]; Kristine B. Gutzkow[40], PhD; Jorunn Evandt[40], Msc; Marina Vafeiadi[41], PhD; Marieke Klein[24,42], PhD; Dennis van der Meer[43,44], PhD; Lannie Ligthart[1], PhD; BIOS Consortium*; Cornelis Kluft[45], PhD; Gareth E. Davies[46], PhD; Christian Hakulinen[15], PhD; Liisa Keltikangas-Järvinen[15], PhD; Barbara Franke[24,47], PhD; Christine M. Freitag[7], PhD; Kerstin Konrad[48,49], PhD; Amaia Hervas[50], PhD, MD; Aranzazu Fernández-Rivas[51], PhD, MD; Agnes Vetro[51], PhD; Olli Raitakari[53,54,55], PhD, MD; Terho Lehtimäki[32], PhD, MD; Robert Vermeiren[4,56], PhD, MD; Timo Strandberg[57], PhD; Katri Räikkönen[15], PhD; Harold Snieder[16], PhD; Stephanie H. Witt[29], PhD; Michael Deuschle[30], MD, Prof; Nancy L. Pedersen[31], PhD; Sara Hägg[31], PhD; Jordi Sunyer[20,21,22,58], PhD, MD; Lude Franke[23], PhD; Jaakko Kaprio[8], PhD, MD; Miina Ollikainen[8], PhD; Terrie E. Moffitt[5,6,36,59], PhD; Henning Tiemeier[10,60], PhD, MD; Marinus H. van Ijzendoorn [61,62], PhD; Caroline Relton[2,3], PhD; Martine Vrijheid[20,21,22], PhD; Sylvain Sebert[26,27,63], PhD; Marjo-Riitta Jarvelin[26,27,64], PhD, MD; Avshalom Caspi[5,6,36,59], PhD; Kathryn L. Evans[13], PhD; Andrew M. McIntosh[12], MD, FRCPsych; Meike Bartels[1], PhD; Dorret Boomsma[1], PhD

1 Department of Biological Psychology, Amsterdam Public Health Research Institute, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
2 Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK.
3 MRC Medical Research Council Integrative Epidemiology Unit, University of Bristol, Bristol, UK.
4 Curium-LUMC, Department of Child and Adolescent Psychiatry, Leiden University Medical Center in Oegstgeest, The Netherlands
5 Department of Psychology and Neuroscience, Duke University, Durham, NC, USA.
6 Center for Genomic and Computational Biology, Duke University, Durham, NC, USA
7 University Hospital Frankfurt; Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy; Frankfurt am Main, Germany
8 Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland
9 Institute of Education and Child Studies, Leiden University, Leiden, the Netherlands.
10 Department of Child and Adolescent Psychiatry/Psychology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands.
11 Generation R Study Group, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands
12 Division of Psychiatry, University of Edinburgh
13 Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh
14 Turku Institute for Advances Studies, University of Turku, Turku Finland
15 Department of Psychology and logopedics, Faculty of Medicine, University of Helsinki, Helsinki, Finland

16 University of Groningen, University Medical Center Groningen, Department of Epidemiology, Groningen, The Netherlands
17 CIBER Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain
18 Dpt. of Biomedical Science, Faculty of Medicine and Health Science, University of Barcelona, Barcelona, Spain
19 Research Group on Statistics, Econometrics ans Health (GRECS), UdG
20 ISGlobal, Barcelona, Spain.
21 Universitat Pompeu Fabra (UPF), Barcelona, Spain.
22 CIBER Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain
23 Department of Genetics, University of Groningen, University Medical Centre Groningen, Groningen, The Netherlands
24 Department of Human Genetics, Donders Institute for Brain, Cognition and Behavior, Radboud University Medical Center, Nijmegen, The Netherlands
25 Department of Psychiatry, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands
26 Center for Life Course Health Research, P.O. Box 5000, 90014, University of Oulu, Oulu, Finland.
27 Biocenter Oulu, University of Oulu, P.O. Box 5000, 90014, Finland
26 Center for Life Course Health Research, P.O. Box 5000, 90014, University of Oulu, Oulu, Finland.
28 Queen's University Belfast
29 Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, Mannheim, Germany
30 Department of Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, Germany
31 The Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Nobels väg 12A, 17165 Stockholm, Sweden
32 Department of Clinical Chemistry, Fimlab Laboratories, and Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere 33520, Finland
33 Department of Special Needs Education, Ghent University, Ghent, Belgium
34 Dunedin Multidisciplinary Health and Development Research Unit, Department of Psychology, University of Otago, Dunedin, New Zealand
35 University of Exeter Medical School, University of Exeter, Exeter, UK
36 Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK
37 Department of Pediatrics, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands
38 Clinical Child & Family Studies, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands
39 Department of Environmental Sciences, Vytautas Magnus University, K. Donelaicio str. 58, 44248 Kaunas, Lithuania
40 Norwegian Institute of Public Health
41 Department of Social Medicine, University of Crete, Greece
42 University Medical Center Utrecht, UMC Utrecht Brain Center, Department of Psychiatry, Utrecht, The Netherlands
43 NORMENT, Division of Mental Health and Addiction, Oslo University Hospital & Institute of Clinical Medicine, University of Oslo, Oslo, Norway.
44 School of Mental Health and Neuroscience, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands
45 Good Biomarker Sciences, Leiden, the Netherlands.
46 Avera Institute for Human Genetics, 3720 W. 69th Street, Sioux Falls, SD, 57108, USA.
47 Department of Psychiatry, Donders Institute for Brain, Cognition and Behavior, Radboud University Medical Center, Nijmegen, The Netherlands
48 University Hospital, RWTH Aachen, Child Neuropsychology Section, Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, Aachen, Germany;
49 JARA-Brain Institute II, Molecular Neuroscience and Neuroimaging (INM-11), RWTH Aachen & Research Centre Juelich, Germany
50 Hospital Universitario Mutua de Terrassa, Child and Adolescent Mental Health Service, Barcelona, Spain
51 Bilbao University Hospital, Psychiatric Service, Osakidetza, Bilbao, Spain
52 Szeged University, Department of Pediatrics and Pediatrics health center, Child and Adolescent Psychiatry, Szeged, Hungary
53 Centre for Population Health Research, University of Turku and Turku University Hospital, Turku, Finland
54 Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland
55 Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland
56 Youz, Parnassia Group, The Hague, the Netherlands
57 Helsinki University Central Hospital, Geriatrics, Helsinki, Finland.
58 IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain
59 Department of Psychiatry and Behavioral Sciences, Duke University School of Medicine, Durham, N.C., U.S.A.

60 Department of Social and Behavioral Science, Harvard TH Chan School of Public Health, Boston USA
61 Department of Psychology, Education and Child Studies, Erasmus University Rotterdam, Rotterdam, the Netherlands.
62 School of Clinical Medicine, University of Cambridge, Cambridge, UK
63 Section of Genomics of Common Disease, Department of Medicine, Imperial College. Paddington, W2 1PG, London, United Kingdom London, Hammersmith Hospital Campus, Burlington Danes Building, Du Cane Road, London, W12 0NN, UK
64 MRC-PHE Centre for Environment and Health, Imperial College London, Praed Street Wing, St Mary's Campus, Paddington, W2 1PG, London, United Kingdom London, Hammersmith Hospital Campus, Burlington Danes Building, Du Cane Road, London, W12 0NN, UK
*Biobank-based Integrative Omics Study Consortium. For a complete list of authors, see the acknowledgements.

Corresponding author: Jenny van Dongen, Ph.D., Department of Biological Psychology, Vrije Universiteit Amsterdam, Van der Boechorststraat 7-9, 2081BT,Amsterdam, Netherlands, tel: +31 (20) 5983570, j.van.dongen@vu.nl .

## Abstract

DNA methylation profiles of aggressive behavior may capture lifetime cumulative effects of genetic, stochastic, and environmental influences associated with aggression. Here, we report the first large meta-analysis of epigenome-wide association studies (EWAS) of aggressive behavior (N=15,324 participants). In peripheral blood samples of 14,434 participants from 18 cohorts with mean ages ranging from 7 to 68 years, 13 methylation sites were significantly associated with aggression (alpha=$1.2 \times 10^{-7}$; Bonferroni correction). In cord blood samples of 2,425 children from five cohorts with aggression assessed at mean ages ranging from 4 to 7 years, 83% of these sites showed the same direction of association with childhood aggression ($r$=0.74, p=0.006) but no epigenome-wide significant sites were found. Top-sites (48 at a false discovery rate of 5% in the peripheral blood meta-analysis or in a combined meta-analysis of peripheral blood and cord blood) have been associated with chemical exposures, smoking, cognition, metabolic traits, and genetic variation (mQTLs). Three genes whose expression levels were associated with top-sites were previously linked to schizophrenia and general risk tolerance. At six CpGs, DNA methylation variation in blood mirrors variation in the brain. On average 44% (range=3-82%) of the aggression–methylation association was explained by current and former smoking and BMI. These findings point at loci that are sensitive to chemical exposures with potential implications for neuronal functions. We hope these results to be a starting point for studies leading to applications as peripheral biomarkers and to reveal causal relationships with aggression and related traits.

## Introduction

Aggression encompasses a range of behaviors, such as bullying, verbal abuse, fighting, and destroying objects. Early life social conditions, including low parental income, separation from a parent, family dysfunction, and maternal smoking during pregnancy are risk factors for childhood aggression[1,2,3]. High levels of aggression are a characteristic of several psychiatric disorders and may also be caused by traumatic brain injury[3], neurodegenerative diseases[4] and alcohol and substance abuse[5,6].

DNA methylation mediates effects of genetic variants in regulatory regions on gene expression[7] and is modifiable by early life social environment, as demonstrated by animal studies[8,9], and by chemical exposures including (prenatal) exposure to cigarette smoke, as illustrated by numerous human studies[10]. Despite the large tissue-specificity of DNA methylation, effects of genetic variants on nearby DNA methylation (*cis* mQTLs) correlate strongly between blood and brain cells[11]. DNA methylation signatures of chemical exposures[12] and maternal rearinging[9] show a certain (but less understood) degree of conservation across tissues.

Large-scale epigenome-wide association studies (EWASs) have become feasible through DNA methylation microarrays applied to blood samples from large cohorts, identifying thousands of loci where methylation in cord blood is associated with maternal smoking[13]. Methylation in blood is associated with depressive symptoms[14] and brain morphology[15], with some evidence for blood DNA methylation signatures being a marker for methylation levels[15] or gene expression[14] in the brain. For several traits, DNA methylation scores based on multiple CpGs from EWAS show better predictive value than currently available polygenic scores[16,17].

Small-scale studies (maximum sample size=260) have provided some evidence that DNA methylation differences in blood, cord blood, and buccal cells are associated with severe forms of aggressive behavior and related problems in children and adults, including (chronic) physical aggression and early onset conduct problems[18–20], but studies on violent aggression in schizophrenia patients (N=134)[21] and a population-based study of continuous aggression symptoms in adults (N=2,029)[22] did not detect epigenome-wide significant sites.

We performed an EWAS meta-analysis of aggressive behavior and closely related constructs. We chose to meta-analyze multiple measures of aggression across ages and sex to maximize sample size. The contribution of genetic influences to aggression is largely stable, at least throughout childhood[23], whereas epigenetic signatures may be dynamic and may differ across cell types and age. Therefore, we performed separate meta-analyses of peripheral blood collected after birth (N=14,434) and cord blood (N=2,425), followed by a

combined meta-analysis (N=15,324) including an examination of heterogeneity of effects. Next, we tested the relationship between aggressive behavior and epigenetic clocks, as associations of lifetime stress[24], exposure to violence[25], and psychiatric disorders[26,27] with accelerated epigenetic ageing have been reported. We performed extensive functional follow-up by integrating our findings with data on gene expression, mQTLs and DNA methylation in brain samples.

## Methods

### Cohorts

Demographic information for the cohorts is provided in **Table 1**. Detailed cohort information is provided in **eAppendix 1**. Informed consent was obtained from all participants. The protocol for each study was approved by the ethical review board of each institution.

### Aggressive behavior

Aggressive behavior was assessed by self-report or reported by parents and teachers. Multiple instruments were used (**eTable 1**): ASEBA Child Behavior Check List (CBCL)[28], Strengths and Difficulties Questionnaire (SDQ) conduct problem scale[29], multidimensional Peer Nomination Inventory (MNPI) aggression scale[30], ASEBA adult self-report (ASR) aggression scale[31], DSM-IV Conduct Disorder Symptom Scale[32], Multidimensional Personality Questionnaire (MPQ) aggression scale[33], and the Hunter-Wolf aggressive behavior scale[34,35]. In four cohorts, a single aggression-related item from personality questionnaires was used. Distributions of aggression scores are provided in **eFigure 1**.

### DNA Methylation BeadChips

DNA methylation was assessed with Illumina BeadChips: the llumina Infinium HumanMethylation450 BeadChip (450k array; majority of cohorts), or the Illumina MethylationEPIC BeadChip (EPIC array). Most cohorts analyzed DNA methylation β-values, which range from 0 to 1, indicating the proportion of DNA that is methylated at a CpG in a sample. Cohort-specific details about DNA methylation profiling, quality control, and normalization are described in **eAppendix 1** and summarized in **eTable 2**.

### Epigenome-wide Association Analysis

EWAS analyses were performed according to a standard operating procedure (http://www.action-euproject.eu/content/data-protocols). In each cohort, the association between DNA methylation level and aggressive behavior was specified under a linear model with DNA methylation as outcome, and correction for relatedness of individuals where applicable. Two models were tested. Model 1 included aggressive behavior, sex, age at blood sampling (not in cohorts with invariable age), white blood cell percentages (measured or imputed), and technical covariates. Model 2 included the same predictors plus body-mass-index (BMI) and smoking status in adolescents and adults (current smoker, former smoker or never smoked). Cohort-specific details and R-code are provided in **eAppendix 1** and **eTable 3,** respectively. The relationship between aggressive behavior and covariates is provided in **eTable 4** based on data from the Netherlands Twin Register (N=2059)**.**

Quality control and filtering of cohort-level EWAS summary statistics is described in **eAppendix 2**. The following probes were removed: on sex chromosomes, methylation sites with more than 5% missing data in a cohort, probes overlapping SNPs affecting the CpG or single base extension site with a minor allele frequency (MAF) > 0.01 in the 1000G EU or GONL population[7], and ambiguous mapping probes reported with an overlap of at least 47 bases per probe[36]. The R package Bacon was used to compute the Bayesian inflation factor and to obtain bias- and inflation-corrected test statistics (**eFigure 2)** prior to meta-analysis[37].

**Meta-analysis**

Fixed-effects meta-analyses were performed in METAL[38]. We used the p-value-based (sample size-weighted) method because the measurement scale of aggressive behavior differs across studies. First, results based on peripheral blood and cord blood data were meta-analyzed separately. Second, a combined meta-analysis was performed of all data. The following cohorts had data available for both cord blood and peripheral blood (from the same children): INMA (which is part of HELIX) and ALSPAC. In the combined meta-analysis, the cord blood data from ALSPAC and INMA were excluded to avoid sample overlap. Statistical significance was assessed considering Bonferroni correction for the number of sites tested (alpha=$1.2 \times 10^{-7}$). Methylation sites that were associated with aggression at the less conservative false discovery rate (FDR) threshold (5%) were included in follow-up analyses. The $I^2$ statistic from METAL was used to describe heterogeneity.

**Follow-up Analyses**

DNA methylation score analyses and epigenetic clock analyses are described in **eAppendix 3** and **eAppendix 4**. Follow-up analyses (**eAppendix 5- eAppendix 10)** were performed on meta-analysis top-sites (FDR<0.05), including a comparison of top-sites with all previously reported associations in the EWAS atlas[39], follow-up analysis of top-sites in two clinical cohorts with blood methylation data (**Table 2**), a cross-tissue analysis (blood, buccal, brain), and association with gene expression level and mQTLs. Analyses of differentially methylated regions (DMRs) are described in **eAppendix 8.** Finally, we performed replication analysis of a previously reported DMR associated with aggression[20] **(eAppendix 9)**.

## Results

### *Peripheral blood meta-analysis*

We performed a meta-analysis of 13 studies with peripheral blood DNA methylation data (N=14,434). The meta-analysis test statistics showed no inflation (**eTable 5, eFigure 3)**. In model 1, methylation at 13 CpGs was associated with aggression (Bonferroni correction; alpha=$1.2x10^{-7}$), and 35 passed a less conservative threshold (FDR 5%; **Figure 1a**). At 28 out of the 35 sites (80%), higher levels of aggression were associated with lower methylation levels. Top-sites showed varying degrees of between-study heterogeneity (mean $I^2$=50%; range=0- 86%, **eTable 6**). Five sites showed significant heterogeneity (alpha=$1.2x10^{-7}$).

### *Cord blood meta-analysis*

The meta-analysis of cord blood (five cohorts; N=2,425) detected no significant CpGs (**eTable 7**). Examining top-sites from the peripheral blood meta-analysis, 12 of the significant, and 33 of the FDR top-sites were assessed in cord blood; 10 (83%), and 25 (71%), respectively, showed the same direction of association (**Figure 1b**). Effect sizes in cord blood correlated significantly with effect sizes in peripheral blood (*r*=0.74, p=0.006 for epigenome-wide significant and *r*=0.51,p=0.003 for FDR top-sites).

### *Combined meta-analysis*

In the combined meta-analysis of peripheral and cord blood data (total sample size=15,324, **eTable 6**), methylation at 13 CpGs was associated with aggression after Bonferroni correction, including ten CpGs from the peripheral blood meta-analysis, and 43 passed a less conservative threshold (FDR 5%, **Table 3**). Among FDR top-sites from both analyses, 13 CpGs were only found in the combined meta-analysis but not in the peripheral blood meta-analysis, while five CpGs from the peripheral blood meta-analysis were no longer significant in the combined meta-analysis (**Figure 1c**).

### *CBCL meta-analysis*

We compared our meta-analysis results to a meta-analysis of cohorts that applied the same aggression instrument; i.e. CBCL (four studies; N=2,286; **Table 1**). No epigenome-wide significant sites were detected (**eFigure 4a**). Examining top-sites from the overall meta-analysis (Model 1), 38 (79%) showed the same direction of association for CBCL aggression in children, and effect sizes correlated strongly (*r*=0.75, p=$6.8x10^{-10}$, **eFigure 4b**).

### *Overlap with CpGs detected in previous EWASs*

We performed enrichment analyses against all previously reported associations with diseases and environmental exposures recorded in the EWAS Atlas[39]. The top ten most

strongly enriched traits are shown in **Figure 1e**. CpGs associated with aggressive behavior showed large overlap with CpGs previously associated with smoking (37 CpGs; corresponding to 77% of aggression-associated CpGs and 0.3% of CpGs that have been previously associated with smoking), and smaller overlap with other smoking traits (e.g. maternal smoking), other chemical exposures (e.g. perinatal exposure to polychlorinated biphenyls (PCBs) and polychlorinated dibenzofurans (PCDFs)). Further overlap includes CpGs associated with alcohol consumption, cognitive function, educational attainment, ageing, and metabolic traits (**eTable 8**).

***Controlling for smoking and BMI***

Model 2 was fitted to test whether the association between DNA methylation and aggressive behavior attenuated after adjusting for the most important postnatal lifestyle factors that influence DNA methylation (smoking and BMI). Examining the 35 CpGs associated with aggression at FDR 5% in peripheral blood, all CpGs showed the same direction of association after adjusting for smoking and BMI (**eTable 6**, **Figure 1d**). Effect sizes were attenuated to varying degrees (mean reduction=44%, range=3-83%). Changes in effect sizes are likely primarily driven by the correction for smoking, since only one top-site has been associated previously with BMI. Some CpGs showed little attenuation, in particular CpGs that have not been previously associated with smoking (e.g.; cg02895948; *PLXN2A,* cg00891184*; KIF1B,* cg1215892*;* intergenic*,* and cg05432213*; ACT1*). In model 2, between-study heterogeneity at top-sites was greatly reduced (adjusted: mean $I^2$=28%, range=0-77%). No CpGs were epigenome-wide significant or FDR-significant in the adjusted meta-analyses.

***DNA methylation scores***

We computed weighted sumscores in NTR (peripheral blood, mean age=36.4, SD=12, N=2,059) based on summary statistics from the peripheral blood meta-analysis without NTR (**Figure 2**). The best score, based on CpGs with $p<1x10^{-3}$ in model 2 (745 CpGs), explained 0.29% of the variance in aggression (p=0.02, not significant after multiple testing correction). This effect was attenuated when age and sex were added to the prediction equation.

***Epigenetic clocks***

Horvath and Hannum epigenetic age acceleration were not associated with aggression (**eTable 9**) in a meta-analysis of 12 studies with peripheral blood DNA methylation data (N=9,554), five studies with cord blood DNA methylation (N=2,225), or in a combined meta-

analysis of 15 studies (N=9,740). There was no significant heterogeneity between cohorts (mean $I^2$=16%, range=0-60%).

***Follow-up in clinical cohorts***

To assess the translation of our observations to aggression-related problem behavior in psychiatric disorders that show comorbidity with aggression, we performed follow-up analyses of top-sites in two clinical cohorts (**Table 2**): the NeuroIMAGE[40] cohort of ADHD cases and controls ($N_{total}$=71) and the FemNAT-CD[41] cohort of female conduct disorder cases and controls ($N_{total}$=100). Results did not replicate (**eAppendix 6, eTable 10, eTable 11, eFigure 5, eFigure 6**).

***Cross-tissue analysis***

To assess the generalizability of our observations in blood to other tissues, we examined the association with CBCL aggression in buccal DNA methylation data (EPIC array), available for 38 top-sites, in children (N=1237) and a child clinical cohort (N=172;**Table 2, eTable 12**) [42]. . We also tested associations with maternal smoking and with child nervous system medication (as indexed by the Anatomical Therapeutic Chemical classification system (ATC N-class))

Correlations between DNA methylation levels in blood and buccal cells, based on 450k data from matched samples (N=22, age=18 years)[43] were available for 36 of these CpGs. The average correlation was weak (*r*=0.25, range=-0.40-0.76). Five CpGs showed a strong correlation between blood and buccal cells (*r*>0.5, **eTable 13**), of which three have been previously associated with (maternal) smoking.

In line with the weak correlation between blood and buccal cell methylation for most top-sites, none of the top-sites was associated with aggression in buccal samples (alpha=0.001, **eTable 14**). Regression coefficients based on analyses in buccal cells and blood overall showed no directional consistency (twin cohort: *r*=0.03, p=0.86; concordant direction: 47%, p=0.87, binomial test, clinical cohort: *r*=0.27, p=0.10; concordant direction: 61%, p=0.26). Exclusion of ancestry outliers did not change these results (**eTable 14**). Of the five CpGs with a large blood-buccal correlation, three showed the same direction of association with aggression in buccal cells from twins, four in clinical cases, and one CpG was nominally associated with aggression in buccal samples from twins; cg11554391 (*AHRR*), $r_{blood\text{-}buccal}$=0.69, $\beta_{aggression}$=-0.0002, p=0.007.

One CpG was significantly associated with maternal smoking in both cohorts: cg04180046 (*MYO1G*), NTR: $\beta_{maternalsmoking=}$0.041, p=6.0x10$^{-6}$, Curium: $\beta_{maternalsmoking=}$0.048, p=7.9x10$^{-5}$ (**eTable 14**). None of the CpGs was associated with medication use (**eTable 14**).

We examined the correlation between DNA methylation levels in blood and brain (N=122)[44] in published DNA methylation data from matched blood samples and four brain regions. Six aggression top-sites (13%) showed significantly correlated DNA methylation levels between blood and one or multiple brain regions: mean *r*=0.52; range=0.45-0.63, alpha=2.6x10$^{-4}$, **eTable 15, eFigure 7**), two of which have not been previously associated with smoking or BMI: cg14560430(*TRIM71),* and cg20673321*(ZNF541).*

**DMRs**

DMR analysis showed that 14 DMPs from our combined meta-analysis reside in regions where multiple correlated methylation sites showed evidence for association with aggressive behavior. DMR analysis also detected additional regions that were not significant in DMP analysis (**eTable 16- eTable 21**). These analyses are described in detail **in eAppendix 8**.

***Replication analysis***

A previous EWAS based on Illumina array data detected a significant DMR in *DRD4* in buccal cells associated with engagement in physical fights[20]. This locus did not replicate in our meta-analyses or in the two cohorts with buccal methylation data (**eTable 22, eAppendix 9)**.

***Gene Expression***

Based on peripheral blood RNA-seq and DNA methylation data (N=2,101)[7], 17 significant DNA methylation-gene expression associations were identified among 15 CpGs and ten transcripts (**Table 3**, **eTable 23**). For most transcripts, a higher methylation level at a CpG site in *cis* correlated with lower expression (82.4%): cg03935116 and *FAM60A*, cg00310412 and *SEMA7A*, cg03707168 and *PPP1R15A,* cg03636183 and *F2RL3,* two intergenic CpGs on chromosome 6, where methylation level correlated negatively with expression levels of *FLOT1*, *TUBB*, *LINC00243,* and six CpGs annotated to *AHRR* were negatively associated with *EXOC3* expression level. Positive correlations were observed between methylation levels at 2 CpGs on chromosome 7 and levels of *RP4-647J21.1* (novel transcript, overlapping *MYO1G)* and between cg02895948 and *PLXNA2.*

***mQTLs***

To gain insight into genetic causes of variation underlying top-sites, we obtained whole-blood mQTL data (N=3,841)[7]. In total, 75 mQTL associations were identified among 34 aggression top-sites (70.8%) and 66 SNPs at the experiment-wide threshold applied by the mQTL study FDR<0.05): 80% were *cis* mQTLs and 20% were *trans* mQTLs (**eTable 24**).

## Discussion

We identified 13 epigenome-wide significant sites (Bonferroni corrected) in the meta-analysis of blood and 13 in the combined meta-analysis of blood and cord blood (16 unique sites). We prioritized 48 top-sites (FDR 5%) for follow-up analyses. Methylation level at three top-sites was associated with expression levels of genes that have been previously linked to psychiatric or behavioral traits in GWASs: *FLOT1* (schizophrenia[45]), *TUBB* (schizophrenia)[45], and *PLXNA2* (general risk tolerance)[46]. Several other loci have functions in the brain and six CpGs showed correlated methylation levels between blood and brain.

The majority of top-sites (77%) were associated with smoking, 46% were associated with maternal smoking, 25% were associated with alcohol consumption, and 15% were associated with perinatal PCB and PCDF exposure. This overlap of aggression top-sites with smoking and other chemical exposures is noteworthy. Methylation levels of top-sites in the Aryl-Hydrocarbon Receptor Repressor gene *AHRR* and several other genes are known to be strongly associated with exposure to cigarette smoke[13,47] and persistent organic pollutants[48]. The best characterized exogenous ligands of the widely expressed Aryl-Hydrocarbon Receptor are environmental contaminants such as benzo[a]pyrene (B[a]P), and TCDD (dioxin), whose neurotoxic and neuro-endocrine effects, including disruption of neuronal proliferation, differentiation, and survival, have been well-characterized[49]. Human prenatal exposure to B[a]P is associated with delayed mental development, lower IQ, anxiety and attention problems[50]. Research on B[a]P neurotoxicity in adults is scarce but a study on coke oven workers found that occupational B[a]P exposure correlates with reduced monoamine, amino acid and choline neurotransmitter levels and with impaired learning and memory[51].

On average 44% (range=3-82%) of the aggression–methylation association was explained by current and former smoking and BMI. Our findings do not merely reflect effects of own smoking: 71% of the top-sites showed the same direction for the prospective association of cord blood methylation at birth and aggression in childhood, and 46% have been associated with maternal prenatal smoking. There is a weak observational association between maternal smoking and child aggression[52]. Our analyses did not adjust for prenatal and postnatal second-hand smoking. Future studies can examine if the link between prenatal maternal smoking and aggression is mediated by DNA methylation.

We found that DNA methylation scores for aggression explained less variation compared to DNA methylation scores for traits such as BMI, smoking and educational attainment. For these traits, EWASs tended to identify more epigenome-wide significant hits[16,17]. The variance in aggression explained by DNA methylation scores was in the same order of magnitude as the variance in height explained by DNA methylation scores (based on

EWASs of height in smaller samples), i.e. less than 1%[16]. More research is needed in particular to delineate if there is a causal link between these methylation sites and aggressive behaviour, since our results may also reflect (residual) confounding by (exposure to second-hand) smoking. One approach to address this could be Mendelian Randomization, in which genetic information (SNPs) is used for causal inference of the effect of an exposure (e.g. DNA methylation) on an outcome (e.g. aggression). This approach previously supported a causal effect of maternal smoking-associated methylation sites in blood on various traits and diseases for which well-powered GWASs have been performed, including schizophrenia[53,54]. For aggressive behavior, the currently available[55] largest GWASs of aggressive behavior included ~16,000 [56] and ~75,000 participants (Ip et al, Multivariate GWA meta-analysis in over 500K observations on aggressive behavior and ADHD symptoms, *submitted*) , respectively. The GWAS by Ip et al detected 3 significant genes in gene-based analysis, but both GWASs did not detect genome-wide significant SNPs and are likely still underpowered. In the future, larger GWASs of aggressive behavior and larger mQTL analyses will allow for powerful Mendelian Randomization for aggression-associated methylation sites.

**Strengths and limitations**

This is the largest EWAS of aggressive behavior to date. The large sample size was achieved by applying a broad phenotype definition, including participants from multiple countries and all ages in a meta-analysis, and analyzing DNA methylation data from blood. A limitation of this approach is that it reduces power to detect age-, sex-, and symptom-specific effects, and that genetic and environmental backgrounds of different populations, as well as non-identical processing methods of methylation data play a role. A limitation of population-based cohorts and even clinical populations is that individuals with extreme levels of aggressive behavior who cause most societal problems are likely underrepresented. Moreover, some studies used measures that tapped features that overlap with but are not necessarily indicative of aggression (e.g. personality traits, anger, oppositional defiant disorder). Future EWASs that specifically focus on more homogeneous aggression measures are therefore warranted. Our meta-analysis approach may identify a common epigenomic signature of aggression-related problems.

Follow-up analysis in independent datasets indicated that these findings do not generalize strongly to buccal cells, and results did not replicate in two clinical cohorts. These were small, used different aggression measures, and one used a different technology (sequencing) in females only.

## Conclusions

We identified associations between aggressive behavior and DNA methylation in blood at CpGs whose methylation level is also associated with exposure to smoking, alcohol consumption, other chemical exposures, and genetic variation. Methylation levels at three top-sites were associated with expression levels of genes that have been previously linked to psychiatric or behavioral traits in GWAS. Our study illustrates both the merit of EWASs based on peripheral tissues to identify environmentally-driven molecular variation associated with behavioral traits and their challenges to tease-out confounders and mediators of the association, and causality. Pursuing full control of potential confounders in behavioral EWAS meta-analyses (including smoking-exposure and other substance-use across the life course, socioeconomic conditions and other, perhaps less obvious, ones) might be unrealistic, and has the potential disadvantage of over-correction. Future studies, including those that integrate EWAS results for multiple traits and exposures, DNA methylation in multiple tissues, and GWASs of multiple traits are warranted to unravel the utility of our results as peripheral biomarkers for pathological mechanisms in other tissues (such as neurotoxicity) and to unravel possible causal relationships with aggression and related traits. We consider this study to be the starting point for such follow-up studies.

## Acknowledgements

This work was supported by ACTION. ACTION receives funding from the European Union Seventh Framework Program (FP7/2007-2013) under grant agreement no 602768. Cohort-specific acknowledgements are provided in **eAppendix 1**. The following authors declare a conflict of interest: Barbara Franke received educational speaking fees from Medice. Andrew M. McIntosh has received research support from Eli Lilly, Janssen and The Sackler Trust and speaker fees from Illumina and Janssen. Christine M. Freitag has received funding by the DFG, BMBF, State of Hessen, and the EU. She receives royalties for books on ASD, ADHD, and MDD.

**BIOS Consortium (Biobank-based Integrative Omics Study)**

**Management Team** Bastiaan T. Heijmans (chair)[1], Peter A.C. 't Hoen[2], Joyce van Meurs[3], Aaron Isaacs[4], Rick Jansen[5], Lude Franke[6].

**Cohort collection** Dorret I. Boomsma[7], René Pool[7], Jenny van Dongen[7], Jouke J. Hottenga[7] (Netherlands Twin Register); Marleen MJ van Greevenbroek[8], Coen D.A. Stehouwer[8], Carla J.H. van der Kallen[8], Casper G. Schalkwijk[8] (Cohort study on Diabetes and Atherosclerosis Maastricht); Cisca Wijmenga[6], Lude Franke[6], Sasha Zhernakova[6], Ettje F. Tigchelaar[6] (LifeLines Deep); P. Eline Slagboom[1], Marian Beekman[1], Joris Deelen[1], Diana van Heemst[9] (Leiden Longevity Study); Jan H. Veldink[10], Leonard H. van den Berg[10] (Prospective ALS Study Netherlands); Cornelia M. van Duijn[4], Bert A. Hofman[11], Aaron Isaacs[4], André G. Uitterlinden[3] (Rotterdam Study).

**Data Generation** Joyce van Meurs (Chair)[3], P. Mila Jhamai[3], Michael Verbiest[3], H. Eka D. Suchiman[1], Marijn Verkerk[3], Ruud van der Breggen[1], Jeroen van Rooij[3], Nico Lakenberg[1].

**Data management and computational infrastructure** Hailiang Mei (Chair)[12], Maarten van Iterson[1], Michiel van Galen[2], Jan Bot[13], Dasha V. Zhernakova[6], Rick Jansen[5], Peter van 't Hof[12], Patrick Deelen[6], Irene Nooren[13], Peter A.C. 't Hoen[2], Bastiaan T. Heijmans[1], Matthijs Moed[1].

**Data Analysis Group** Lude Franke (Co-Chair)[6], Martijn Vermaat[2], Dasha V. Zhernakova[6], René Luijk[1], Marc Jan Bonder[6], Maarten van Iterson[1], Patrick Deelen[6], Freerk van Dijk[14], Michiel van Galen[2], Wibowo Arindrarto[12], Szymon M. Kielbasa[15], Morris A. Swertz[14], Erik. W van Zwet[15], Rick Jansen[5], Peter-Bram 't Hoen (Co-Chair)[2], Bastiaan T. Heijmans (Co-Chair)[1].

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. Genomics 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

**Figure 1** DNA methylation associated with aggressive behavior in a large blood-based meta-analysis

a) Manhattan plot showing the fixed effects meta-analysis p-values for the association between aggressive behavior and DNA methylation level based on the meta-analysis of peripheral blood. The blue horizontal line denotes the FDR-threshold (5%) and the red line indicates the Bonferroni threshold. b) Effects sizes of top-sites from the meta-analysis of aggression in peripheral blood (x-axis) versus effects sizes from the meta-analysis of aggression in cord blood (y-axis). c) Venn diagram showing the numbers and overlap of CpGs detected at FDR 5% in the meta-analysis of peripheral blood and the combined meta-analysis and cord blood and peripheral blood. d) Effects sizes of top-sites from the meta-analysis of aggression in peripheral blood model 1 (x-axis) versus effects sizes from the meta-analysis of aggression in peripheral blood model 2; adjusted for smoking and BMI (y-axis). e) Top enriched traits based on enrichment analysis with all 48 top-sites. The third column shows how many of the 48 CpGs have been previously associated with the trait in the first column. The last column shows the overlap as a percentage of the total number of CpGs previously associated with the trait in column 1 (e.g. 0.3% of all CpGs previously associated with smoking are also associated with aggression in the current meta-analysis. d) In panel b and d, CpGs that have not been previously associated with smoking in the meta-analysis by Joehanes et al[47] are plotted in red.

**Figure 2** Prediction of aggression by DNA methylation scores

The bars indicate how much of the variance in ASEBA adult self-report (ASR) aggression scores were explained by DNA methylation scores in NTR (N=2059, peripheral blood, 450k array). Scores were created based on weights from the peripheral blood meta-analysis with NTR excluded (N=12,375). The y-axis shows percentage of variance explained. Different colors denote DNA methylation scores created with different numbers of CpGs that were selected on their p-value in the meta-analysis (see legend). From left to right, the first three plots show DNA methylation scores created based on weights obtained from the meta-analysis of EWAS model 1, and plots 4 till 6 show DNA methylation scores created based on weights obtained from the meta-analysis of EWAS model 2. Each DNA methylation score was tested for association with aggression in three model: the simplest model (first plot) included aggression as outcome variable, and DNA methylation score as predicted plus technical covariates and cell counts. The second model additionally included sex and age as predictors. The third model additionally included sex, age, and smoking as predictors. Stars denote nominal p-values < 0.05 (not corrected for multiple testing).

**Table 1** Discovery cohorts

| Cohort | N, M1 | N, M2 | % female | % current smoker | % former smoker | DNA age, Mean (SD), y[e] | Aggression survey | Array | Aggression, Mean (SD) | Time between survey and DNA, Mean (min, max), y[f] |
|---|---|---|---|---|---|---|---|---|---|---|
| Peripheral blood | | | | | | | | | | |
| ALSPAC[57] | 865 | 865 | 49.4 | 0 | 0 | 7.5 (0.2) | SDQ[29] | 450k | 1.5 (1.4) | 0.7 (0.0, 2.1) |
| Dunedin[58] | 767 | 764 | 46.3 | 33.8 | 13.7 | 26.0 (0) | MPQ[33] | 450k | 23.3 (19.3) | 0 |
| E-Risk[59] | 1629 | 1601 | 49.8 | 22.7 | 0 | 18.0 (0) | DSM-IV Conduct Disorder[32] | 450k | 2.2 (2.3) | 0 |
| FinnTwin12[60] | 757 | 757 | 59.2 | 46.0* | NA | 22.4 (0.7) | MNPI[30] | 450k | 0.6 (0.7) | 10.4 (9.0,13.0) |
| GS:SFHS[61] | 4609 | 4421 | 67.9 | 18.9 | 29.5 | 46.6 (14.0) | 1 item, from GHQ 28[62a] | EPIC | 0.1(0.3) | 0 |
| GLAKU[63] | 192 | 177 | 56.3 | 1.7 | 0 | 12.3 (0.5) | CBCL[28] | EPIC | 3.9 (3.8) | 0 |
| HELIX[64] | 1058 | 1058 | 44.9 | NA | NA | 8.0 (1.6) | CBCL[28] | 450k | 5.2 (5.0) | 0 |
| LLD[65] | 683 | 683 | 59.4 | 19.0 | 33.1 | 43.9 (11.6) | 1 item, personality questionnaire[b] | 450k | 1.9 (0.9) | 0.1, (0.0,0.3) |
| NFBC1966[66] | 740 | 740 | 56.9 | 29.9 | 23.8 | 31.0 (0) | 1 item, from TCI-NS4[c] | 450k | 0.8 (0.4) | 0.6 (0.0,10) |
| NFBC1986[66] | 517 | 517 | 53.8 | 36.7 | 41.9 | 16.0 (0) | ASR[31] | 450k | 4.3 (2.6) | 0.6 (0.0,10) |
| NTR[67] | 2059 | 2049 | 69.2 | 18.3 | 22.5 | 36.4 (12.0) | ASR[31] | 450k | 2.8 (3.1) | -2.6 (-10.0, 8.0) |
| SATSA[68] | 377 | 377 | 60.2 | 17.0 | 4.0 | 70.2 (9.7) | 1 item, from EAS[69,70d] | 450k | 2.0 (1.07) | -2.0 (-10.0,5.0) |
| YFS[71] | 181 | 181 | 63.0 | 30.9 | 27.5 | 19.2 (3.3) | Hunter-Wolf[34,35] | 450k | 3.5 (0.9) | 0 |
| Cord blood | | | | | | | | | | |
| ALSPAC[57] | 808 | 808 | 50.4 | 0 | 0 | 0 (0) | SDQ[29] | 450k | 1.5 (1.4) | -6.8 (-6.8,-6.8) |
| GECKO[72] | 196 | 186 | 51.5 | 0 | 0 | 0 (0) | SDQ[29] | 450k | 1.1 (1.4) | -5.9(-5.1,- 6.9) |
| Generation R[73] | 806 | 718 | 49.4 | 0 | 0 | 0 (0) | CBCL[28] | 450k | 5.2 (5.1) | -5.9 (-5.2, -8.3) |
| INMA[74] | 385 | 385 | 48.8 | 0 | 0 | 0 (0) | SDQ[29] | 450k | 1.8 (1.7) | -6,9 (-8,3,-6,2) |
| Poseidon[75] | 230 | 230 | 54.3 | 0 | 0 | 0 (0) | CBCL[28] | 450k | 9.4 (5.9) | -3.8 (-3.6, -4) |

ALSPAC=Avon Longitudinal Study of Parents and Children, Dunedin= Dunedin Multidisciplinary Health and Development Study, E-Risk= E-Risk Twin Study, FinnTwin12=Finnish Twin Cohort, GS:SFHS= Generation Scotland: Scottish Family Health Study, GLAKU= Glycyrrhizin in Licorice cohort,HELIX=The Human Early-Life Exposome, LLD= LifeLines-DEEP, NFBC1966=Northern Finland Birth Cohort 1966, NFBC1986= Northern Finland Birth Cohort 1986, NTR= Netherlands Twin Register, SATSA= Swedish Adoption/Twin Study of Aging, YFS= Young Finns Study, GECKO= Groningen Expert Center for Kids with Obesity, Generation R=Generation R Study, INMA= The INMA-INfancia y Medio Ambiente (Environment and Childhood) Project. Poseidon= Pre-, peri- and postnatal Stress in human and non-human offspring: A translational approach to study Epigenetic Impact on DepressiON. SDQ= Strengths and Difficulties Questionnaire (SDQ), conduct problems. MPQ= Multidimensional Personality Questionnaire aggression. DSM-IV Conduct Disorder =DSM-IV Conduct Disorder Symptom Scale. MNPI= Multidimensional Peer Nomination Inventory, aggression. CBCL= Child Behavior Checklist, Aggressive Behavior scale. GHQ= General Health Questionnaire. TCI-NS4=Temperament and Character Inventory- Novelty Seeking. ASR=Adult self-report, aggression scale. EAS= Emotionality, activity, sociability scale. Hunter-Wolf= Hunter-Wolf aggressive behavior scale. [a]Have you recently been getting edgy and bad-tempered? [b]Could you indicate to what

extent the following statement applies to you? I am known for being short-tempered and irritable? [c]I lose my temper more quickly than most people. [d]People think I am hot-tempered an temperamental. [e]Age at DNA sample collection. [f]Time between DNA sample collection and phenotype measure: DNA minus phenotype. M1=model1. M2=model2. Model 1 included the following predictors: aggressive behavior, sex, age at blood sampling (if applicable), white blood cell percentages (measured or imputed), and technical covariates. Model 2 included the same predictors as model 1 plus BMI and smoking status (smoking status was not included in model 2 in cohorts that assessed DNA methylation in children). *The percentage shows current and former smokers combined. NA=not assessed.

**Table 2** Follow-up cohorts

| Cohort | Type | DNA methylation | Phenotype | N | % female | Mean age (SD) | Aggression mean (SD) |
|---|---|---|---|---|---|---|---|
| NeuroIMAGE[40] | Clinical cohort; ADHD | Illumina EPIC | Callous Traits | 71 | 28.2 | 21 (2.9) | 9.3 (4.4) |
| FemNAT-CD[41] | Clinical cohort; Conduct disorder | HpaII methylation Sequencing | Case-control status | Total: 100 Cases: 50 Controls:50 | 100 | Cases: 16.1(1.6) Controls: 15.8(1.5) | NA |
| ACTION – NTR[42] | Twin cohort, selected on aggression (high-low) | Illumina EPIC | CBCL aggression | 1237 | 47.4 | 9.6 (1.9) | 5.0 (5.4) |
| ACTION-Curium-LUMC[42] | Clinical cohort; children with severe and complex mental health problems | Illumina EPIC | CBCL aggression | 172 | 25.6 | 9.6 (1.7) | 13.1 (7.6) |

NeuroIMAGE=The NeuroIMAGE study is a follow-up of the Dutch part of the International Multicenter ADHD Genetics (IMAGE) project. FemNAT-CD= Neurobiology and Treatment of Adolescent Female Conduct Disorder. ACTION= Aggression in Children: Unraveling gene-environment interplay to inform Treatment and InterventiON strategies. NTR= Netherlands Twin Register.

**Table 3** Top-sites associated with aggressive behavior from the combined EWAMA of cord blood and peripheral blood (FDR 5%)

| CpG ID | CHR | Position* | *Gene* | *Gene Expression Associated With CpGs* | N M1 | Zscore M1 | P value M1 | Zscore M2 | P value M2 |
|---|---|---|---|---|---|---|---|---|---|
| cg05575921 | 5 | 373378 | *AHRR* | *EXOC3* | 15666 | -8.995 | 2.36E-19 | -4.159 | 3.20E-05 |
| cg21161138 | 5 | 399360 | *AHRR* | *EXOC3* | 15661 | -7.573 | 3.66E-14 | -3.155 | 1.61E-03 |
| cg26703534 | 5 | 377358 | *AHRR* | *EXOC3* | 15665 | -6.695 | 2.16E-11 | -2.058 | 3.96E-02 |
| cg14753356 | 6 | 30720108 | | *FLOT1* | 15666 | -6.672 | 2.52E-11 | -3.342 | 8.33E-04 |
| cg22132788 | 7 | 45002486 | *MYO1G* | | 10847 | 6.313 | 2.74E-10 | 3.637 | 2.76E-04 |
| cg06126421 | 6 | 30720080 | | *FLOT1, TUBB, LINC00243* | 10864 | -6.196 | 5.78E-10 | -2.154 | 3.13E-02 |
| cg07826859 | 7 | 45020086 | *MYO1G* | | 10863 | -6.017 | 1.77E-09 | -3.665 | 2.48E-04 |
| cg09935388 | 1 | 92947588 | *GFI1* | | 15661 | -5.906 | 3.51E-09 | -3.222 | 1.27E-03 |
| cg25648203 | 5 | 395444 | *AHRR* | *EXOC3* | 15657 | -5.583 | 2.37E-08 | -2.233 | 2.55E-02 |
| cg12062133 | 8 | 142548839 | | | 14482 | 5.462 | 4.71E-08 | 4.881 | 1.06E-06 |
| cg05951221 | 2 | 233284402 | | | 10864 | -5.443 | 5.25E-08 | -1.679 | 9.32E-02 |
| cg14817490 | 5 | 392920 | *AHRR* | *EXOC3* | 10863 | -5.407 | 6.43E-08 | -2.152 | 3.14E-02 |
| cg14179389 | 1 | 92947961 | *GFI11* | | 15666 | -5.35 | 8.80E-08 | -3.888 | 1.01E-04 |
| cg05432213 | 15 | 35086985 | *ACTC1* | | 15666 | 5.144 | 2.68E-07 | 4.87 | 1.12E-06 |
| cg03636183 | 19 | 17000585 | *F2RL3* | *F2RL3* | 15666 | -5.124 | 3.00E-07 | -0.909 | 3.63E-01 |
| cg09022230 | 7 | 5457225 | *TNRC18* | | 15666 | -5.071 | 3.95E-07 | -3.024 | 2.49E-03 |
| cg12803068 | 7 | 45002919 | *MYO1G* | *RP4-647J21.1* | 15666 | 4.93 | 8.22E-07 | 2.493 | 1.27E-02 |
| cg23916896 | 5 | 368804 | *AHRR* | | 15652 | -4.915 | 8.86E-07 | -2.332 | 1.97E-02 |
| cg04180046 | 7 | 45002736 | *MYO1G* | *RP4-647J21.1* | 15665 | 4.884 | 1.04E-06 | 2.989 | 2.80E-03 |
| cg02228160 | 5 | 143192067 | *HMHB1* | | 10852 | 4.867 | 1.13E-06 | 3.451 | 5.58E-04 |
| cg03519879 | 14 | 74227499 | *C14orf43* | | 15663 | -4.859 | 1.18E-06 | -3.609 | 3.08E-04 |
| cg00310412 | 15 | 74724918 | *SEMA7A* | *SEMA7A* | 15666 | -4.854 | 1.21E-06 | -2.608 | 9.11E-03 |
| cg13165240 | 17 | 3715743 | *C17orf85* | | 15664 | 4.838 | 1.31E-06 | 4.436 | 9.16E-06 |
| cg02895948 | 1 | 208204062 | *PLXNA2* | *PLXNA2* | 10865 | -4.811 | 1.51E-06 | -4.448 | 8.68E-06 |
| cg12147622 | 10 | 74021432 | | | 15662 | -4.796 | 1.62E-06 | -3.312 | 9.26E-04 |
| cg26883434 | 5 | 111091560 | *C5orf13* | | 14540 | 4.773 | 1.81E-06 | 4.739 | 2.15E-06 |
| cg03991871 | 5 | 368447 | *AHRR* | *EXOC3* | 10857 | -4.753 | 2.01E-06 | -2.374 | 1.76E-02 |
| cg06946797 | 16 | 11422409 | | | 15666 | -4.75 | 2.03E-06 | -3.317 | 9.08E-04 |
| cg00891184 | 1 | 10272185 | *KIF1B* | | 15662 | 4.746 | 2.07E-06 | 4.421 | 9.82E-06 |
| cg09243533 | 1 | 19281949 | *IFFO2* | | 15666 | -4.74 | 2.14E-06 | -4.003 | 6.26E-05 |
| cg03935116 | 12 | 31476565 | *FAM60A* | *FAM60A* | 15665 | -4.735 | 2.19E-06 | -3.664 | 2.48E-04 |
| cg11554391 | 5 | 321320 | *AHRR* | | 15666 | -4.717 | 2.39E-06 | -2.731 | 6.32E-03 |
| cg19825437 | 3 | 169383292 | | | 15664 | -4.663 | 3.12E-06 | -3.094 | 1.98E-03 |
| cg00624037 | 12 | 89315201 | | | 15663 | 4.633 | 3.61E-06 | 4.081 | 4.49E-05 |
| cg01940273 | 2 | 233284934 | | | 15666 | -4.621 | 3.82E-06 | -0.305 | 7.61E-01 |
| cg25949550 | 7 | 145814306 | *CNTNAP2* | | 15666 | -4.615 | 3.94E-06 | -2.333 | 1.96E-02 |
| cg23067299 | 5 | 323907 | *AHRR* | | 10865 | 4.615 | 3.94E-06 | 3.21 | 1.33E-03 |
| cg04387347 | 16 | 88537187 | *ZFPM1* | | 9563 | 4.603 | 4.17E-06 | 2.678 | 7.42E-03 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| cg02325250 | 5 | 131409289 | *CSF2* | | 15664 | -4.597 | 4.28E-06 | -3.635 | 2.78E-04 |
| cg14560430 | 3 | 32863175 | *TRIM71* | | 15665 | -4.569 | 4.90E-06 | -3.924 | 8.70E-05 |
| cg03844894 | 15 | 35086967 | *ACTC1* | | 15666 | 4.567 | 4.94E-06 | 4.176 | 2.97E-05 |
| cg21611682 | 11 | 68138269 | *LRP5* | | 14859 | -4.561 | 5.08E-06 | -1.721 | 8.53E-02 |
| cg20673321 | 19 | 48049233 | *ZNF541* | | 15666 | 4.538 | 5.67E-06 | 4.672 | 2.98E-06 |

*Genome build 37. M1=Model 1. M2=Model 2

## References


1. Tremblay, R. E. Physical Aggression During Early Childhood: Trajectories and Predictors. *Pediatrics* (2004). doi:10.1542/peds.114.1.e43
2. Tremblay, R. E., Vitaro, F. & Côté, S. M. Developmental Origins of Chronic Physical Aggression: A Bio-Psycho-Social Model for the Next Generation of Preventive Interventions. *Annu. Rev. Psychol.* (2017). doi:10.1146/annurev-psych-010416-044030
3. Tateno, A., Jorge, R. E. & Robinson, R. G. Clinical Correlates of Aggressive Behavior After Traumatic Brain Injury. *J. Neuropsychiatry Clin. Neurosci.* (2014). doi:10.1176/jnp.15.2.155
4. Volicer, L. & Hurley, A. C. Review Article: Management of Behavioral Symptoms in Progressive Degenerative Dementias. *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* (2003). doi:10.1093/gerona/58.9.m837
5. Moore, T. M. *et al.* Drug abuse and aggression between intimate partners: A meta-analytic review. *Clinical Psychology Review* (2008). doi:10.1016/j.cpr.2007.05.003
6. Boles, S. M. & Miotto, K. Substance abuse and violence: A review of the literature. *Aggression and Violent Behavior* (2003). doi:10.1016/S1359-1789(01)00057-X
7. Bonder, M. J. *et al.* Disease variants alter transcription factor levels and methylation of their binding sites. *Nat. Genet.* **49,** 131–138 (2017).
8. Weaver, I. C. G. *et al.* Epigenetic programming by maternal behavior. *Nat. Neurosci.* **7,** 847–54 (2004).
9. Provencal, N. *et al.* The Signature of Maternal Rearing in the Methylome in Rhesus Macaque Prefrontal Cortex and T Cells. *J. Neurosci.* (2012). doi:10.1523/jneurosci.1470-12.2012
10. Martin, E. M. & Fry, R. C. Environmental Influences on the Epigenome: Exposure-Associated DNA Methylation in Human Populations. *Annu. Rev. Public Health* (2018). doi:10.1146/annurev-publhealth-040617-014629
11. Qi, T. *et al.* Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. *Nat. Commun.* (2018). doi:10.1038/s41467-018-04558-1
12. Tsai, P. C. *et al.* Smoking induces coordinated DNA methylation and gene expression changes in adipose tissue with consequences for metabolic health. *Clin. Epigenetics* (2018). doi:10.1186/s13148-018-0558-0
13. Joubert, B. R. *et al.* DNA Methylation in Newborns and Maternal Smoking in Pregnancy: Genome-wide Consortium Meta-analysis. *Am. J. Hum. Genet.* **98,** 680–696 (2016).
14. Jovanova, O. S. *et al.* DNA methylation signatures of depressive symptoms in middle-

aged and elderly persons: Meta-analysis of multiethnic epigenome-wide studies. *JAMA Psychiatry* (2018). doi:10.1001/jamapsychiatry.2018.1725

15. Jia, T. *et al.* Epigenome-wide meta-analysis of blood DNA methylation and its association with subcortical volumes : findings from the ENIGMA Epigenetics Working. *bioRchiv* (2018). doi:https://doi.org/10.1101/460444
16. Shah, S. *et al.* Improving Phenotypic Prediction by Combining Genetic and Epigenetic Associations. *Am. J. Hum. Genet.* (2015). doi:10.1016/j.ajhg.2015.05.014
17. McCartney, D. L. *et al.* Epigenetic prediction of complex traits and death. *Genome Biol.* (2018). doi:10.1186/s13059-018-1514-1
18. Guillemin, C. *et al.* DNA methylation signature of childhood chronic physical aggression in T cells of both men and women. *PLoS One* (2014). doi:10.1371/journal.pone.0086822
19. Cecil, C. A. M. *et al.* Neonatal DNA methylation and early-onset conduct problems: A genome-wide, prospective study. *Dev. Psychopathol.* (2018). doi:10.1017/S095457941700092X
20. Cecil, C. A. M. *et al.* DRD4 methylation as a potential biomarker for physical aggression: An epigenome-wide, cross-tissue investigation. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* (2018). doi:10.1002/ajmg.b.32689
21. Mitjans, M. *et al.* Violent aggression predicted by multiple pre-adult environmental hits. *Mol. Psychiatry* (2018). doi:10.1038/s41380-018-0043-3
22. van Dongen, J. *et al.* Epigenome-wide association study of aggressive behavior. *Twin Res. Hum. Genet.* **18,** 686–698 (2015).
23. Lubke, G. H., McArtor, D. B., Boomsma, D. I. & Bartels, M. Genetic and environmental contributions to the development of childhood aggression. *Dev. Psychol.* **54,** 39–50 (2018).
24. Zannas, A. S. *et al.* Lifetime stress accelerates epigenetic aging in an urban, African American cohort: relevance of glucocorticoid signaling. *Genome Biol.* **16,** 266 (2015).
25. Jovanovic, T. *et al.* Exposure to Violence Accelerates Epigenetic Aging in Children. *Sci. Rep.* **7,** (2017).
26. Han, L. K. M. *et al.* Epigenetic Aging in Major Depressive Disorder. *Am. J. Psychiatry* appi.ajp.2018.1 (2018). doi:10.1176/appi.ajp.2018.17060595
27. Ori, A. P. S. *et al.* Schizophrenia is characterized by age- and sex-specific effects on epigenetic aging. *bioRxiv* 727859 (2019). doi:10.1101/727859
28. Thomas M Achenbach, C. E. Manual for the Child Behavior Checklist. *Burlingt.* **7,** (1991).
29. Goodman, R. The Strengths and Difficulties Questionnaire: a research note. *J. Child Psychol. Psychiatry* **38,** 581–6 (1997).

30. Pulkkinen, L., Kaprio, J. & Rose, R. J. Peers, teachers and parents as assessors of the behavioural and emotional problems of twins and their adjustment: the Multidimensional Peer Nomination Inventory. *Twin Res.* **2,** 274–285 (1999).
31. Achenbach, T. M. & Rescorla, L. a. Manual for the ASEBA Adult Forms & Profiles. *English* University of Vermont, Research Center for Childre (2003). doi:10.1017/CBO9781107415324.004
32. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, (DSM IV). *Washingt. DC, APA* **Fourth Ed.,** 915 (1994).
33. Tellegen, A. *et al.* Personality Similarity in Twins Reared Apart and Together. *J. Pers. Soc. Psychol.* **54,** 1031–1039 (1988).
34. Wolf, T. M., Sklov, M. C., Wenzl, P. A., Hunter, S. M. & Berenson, G. S. Validation of a measure of type A behavior pattern in children: Bogalusa heart study. *Child Dev.* **53,** 126–135 (1982).
35. Ravaja, N., Keltikangas-Järvinen, L. & Keskivaara, P. Type A factors as predictors of changes in the metabolic syndrome precursors in adolescents and young adults--a 3-year follow-up study. *Health Psychol.* **15,** 18–29 (1996).
36. Chen, Y. *et al.* Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics* **2294,** (2013).
37. Iterson, M. Van, Zwet, E. W. Van & Heijmans, B. T. Controlling bias and inflation in association studies using the empirical null distribution. *Genome Biol.* 1–13 (2017). doi:10.1186/s13059-016-1131-9
38. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* (2010). doi:10.1093/bioinformatics/btq340
39. Li, M. *et al.* EWAS Atlas: A curated knowledgebase of epigenome-wide association studies. *Nucleic Acids Res.* (2019). doi:10.1093/nar/gky1027
40. von Rhein, D. *et al.* The NeuroIMAGE study: a prospective phenotypic, cognitive, genetic and MRI study in children with attention-deficit/hyperactivity disorder. Design and descriptives. *Eur. Child Adolesc. Psychiatry* (2015). doi:10.1007/s00787-014-0573-4
41. Freitag, C. M. *et al.* Conduct disorder in adolescent females: current state of research and study design of the FemNAT-CD consortium. *European Child and Adolescent Psychiatry* (2018). doi:10.1007/s00787-018-1172-6
42. Hagenbeek, F. A. *et al.* Urinary Amine and Organic Acid Metabolites Evaluated as Markers for Childhood Aggression: The ACTION Biomarker Study. *Front. Psychiatry* **11,** (2020).
43. van Dongen, J. *et al.* Genetic and environmental influences interact with age and sex

in shaping the human methylome. *Nat. Commun.* **7,** 11115 (2016).

44. Hannon, E., Lunnon, K., Schalkwyk, L. & Mill, J. Interindividual methylomic variation across blood, cortex, and cerebellum: Implications for epigenetic studies of neurological and neuropsychiatric phenotypes. *Epigenetics* **10,** 1024–1032 (2015).
45. Ripke, S. *et al.* Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511,** 421–427 (2014).
46. Karlsson Linnér, R. *et al.* Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nat. Genet.* (2019). doi:10.1038/s41588-018-0309-3
47. Joehanes, R. *et al.* Epigenetic Signatures of Cigarette Smoking. *Circ. Cardiovasc. Genet.* **9,** 436–447 (2016).
48. Su, K. Y. *et al.* Perinatal polychlorinated biphenyls and polychlorinated dibenzofurans exposure are associated with DNA methylation changes lasting to early adulthood: Findings from Yucheng second generation. *Environ. Res.* (2019). doi:10.1016/j.envres.2019.01.001
49. Juricek, L. & Coumoul, X. The aryl hydrocarbon receptor and the nervous system. *International Journal of Molecular Sciences* (2018). doi:10.3390/ijms19092504
50. Chepelev, N. L., Moffat, I. D., Bowers, W. J. & Yauk, C. L. Neurotoxicity may be an overlooked consequence of benzo[a]pyrene exposure that is relevant to human health risk assessment. *Mutation Research - Reviews in Mutation Research* **764,** 64–89 (2015).
51. Niu, Q., Zhang, H., Li, X. & Li, M. Benzo[a]pyrene-induced neurobehavioral function and neurotransmitter alterations in coke oven workers. *Occup. Environ. Med.* **67,** 444–448 (2010).
52. Malanchini, M. *et al.* Aggressive behaviour in childhood and adolescence: The role of smoking during pregnancy, evidence from four twin cohorts in the EU-ACTION consortium. *Psychol. Med.* (2019). doi:10.1017/S0033291718001344
53. Richardson, T. G. *et al.* An integrative approach to detect epigenetic mechanisms that putatively mediate the influence of lifestyle exposures on disease susceptibility. *Int. J. Epidemiol.* **48,** 887–898 (2019).
54. Wiklund, P. *et al.* DNA methylation links prenatal smoking exposure to later life health outcomes in offspring. *Clin. Epigenetics* **11,** (2019).
55. Odintsova, V. V. *et al.* GENOMICS OF HUMAN AGGRESSION: current state of genome-wide studies and an automated systematic review tool. *Psychiatr. Genet.* **29,** 170–190 (2019).
56. Tielbeek, J. J. *et al.* Genome-Wide Association Studies of a Broad Spectrum of Antisocial Behavior. *JAMA Psychiatry* **74,** 1242 (2017).

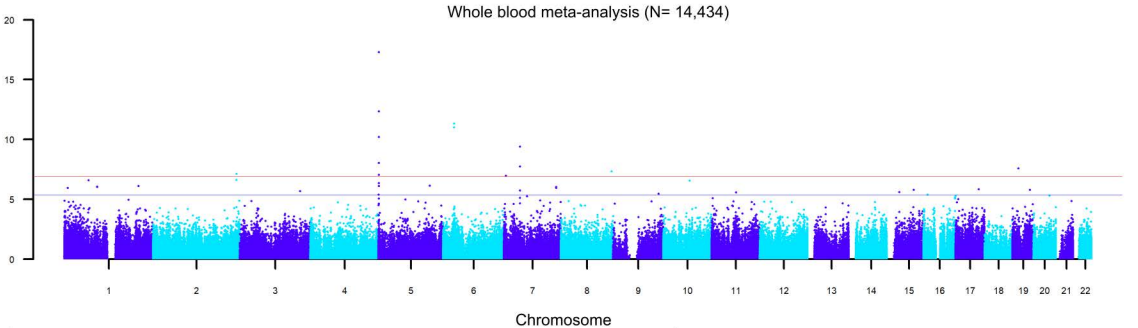
57. Relton, C. L. *et al.* Data Resource Profile: Accessible Resource for Integrated Epigenomic Studies (ARIES). *Int. J. Epidemiol.* (2015). doi:10.1093/ije/dyv072
58. Poulton, R., Moffitt, T. E. & Silva, P. A. The Dunedin Multidisciplinary Health and Development Study: overview of the first 40 years, with an eye to the future. *Social Psychiatry and Psychiatric Epidemiology* **50,** 679–693 (2015).
59. Moffitt, T. E. *et al.* Teen-aged mothers in contemporary Britain. *J. Child Psychol. Psychiatry Allied Discip.* **43,** 727–742 (2002).
60. Kaprio, J. The Finnish Twin Cohort Study: An update. *Twin Res. Hum. Genet.* **16,** 157–162 (2013).
61. Smith, B. H. *et al.* Cohort profile: Generation scotland: Scottish family health study (GS: SFHS). The study, its participants and their potential for genetic research on health and illness. *Int. J. Epidemiol.* (2013). doi:10.1093/ije/dys084
62. Goldberg, D. P. & Hillier, V. F. A scaled version of the General Health Questionnaire. *Psychol. Med.* (1979). doi:10.1017/S0033291700021644
63. Strandberg, T. E., Järvenpää, A. L., Vanhanen, H. & McKeigue, P. M. Birth outcome in relation to licorice consumption during pregnancy. *Am. J. Epidemiol.* (2001). doi:10.1093/aje/153.11.1085
64. Vrijheid, M. *et al.* The human early-life exposome (HELIX): Project rationale and design. *Environmental Health Perspectives* **122,** 535–544 (2014).
65. Tigchelaar, E. F. *et al.* Cohort profile: LifeLines DEEP, a prospective, general population cohort study in the northern Netherlands: study design and baseline characteristics. *BMJ Open* **5,** e006772 (2015).
66. Rantakallio, P. The longitudinal study of the northern Finland birth cohort of 1966. *Paediatr. Perinat. Epidemiol.* **2,** 59–88 (1988).
67. Boomsma, D. I. *et al.* Netherlands Twin Register: From Twins to Twin Families. *Twin Res. Hum. Genet.* **9,** 849–857 (2006).
68. Pedersen, N. L. *et al.* The Swedish Adoption Twin Study of Aging: An update. *Acta Genet. Med. Gemellol. (Roma).* **40,** 7–20 (1991).
69. Buss, A. H. & Plomin, R. *Temperament Early developing personality traitsle*. (Lawrence Erlbaum Associates Inc, 1984).
70. Plomin, R., Pedersen, N. L., McClearn, G. E., Nesselroade, J. R. & Bergeman, C. S. EAS temperaments during the last half of the life span: twins reared apart and twins reared together. *Psychol. Aging* **3,** 43–50 (1988).
71. Raitakari, O. T. *et al.* Cohort profile: The cardiovascular risk in young Finns study. *Int. J. Epidemiol.* **37,** 1220–1226 (2008).
72. L'Abée, C. *et al.* Cohort Profile: The GECKO Drenthe study, overweight programming during early childhood. *Int. J. Epidemiol.* **37,** 486–489 (2008).

73. Kruithof, C. J. *et al.* The Generation R Study: Biobank update 2015. *Eur. J. Epidemiol.* **29,** 911–927 (2014).

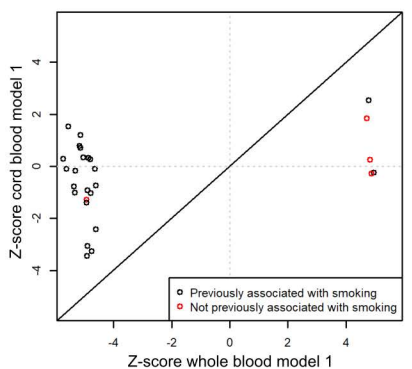
74. Guxens, M. *et al.* Cohort profile: The INMA-INfancia y Medio Ambiente-(environment and childhood) project. *Int. J. Epidemiol.* **41,** 930–940 (2012).

75. Witt, S. H. *et al.* Impact on birth weight of maternal smoking throughout pregnancy mediated by DNA methylation. *BMC Genomics* (2018). doi:10.1186/s12864-018-4652-7

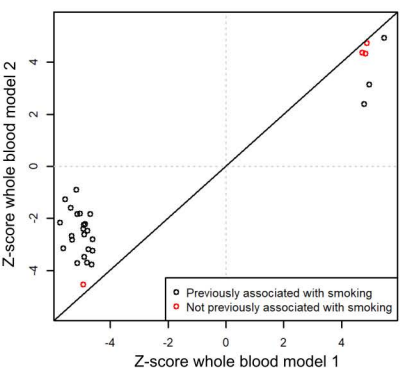
a




b



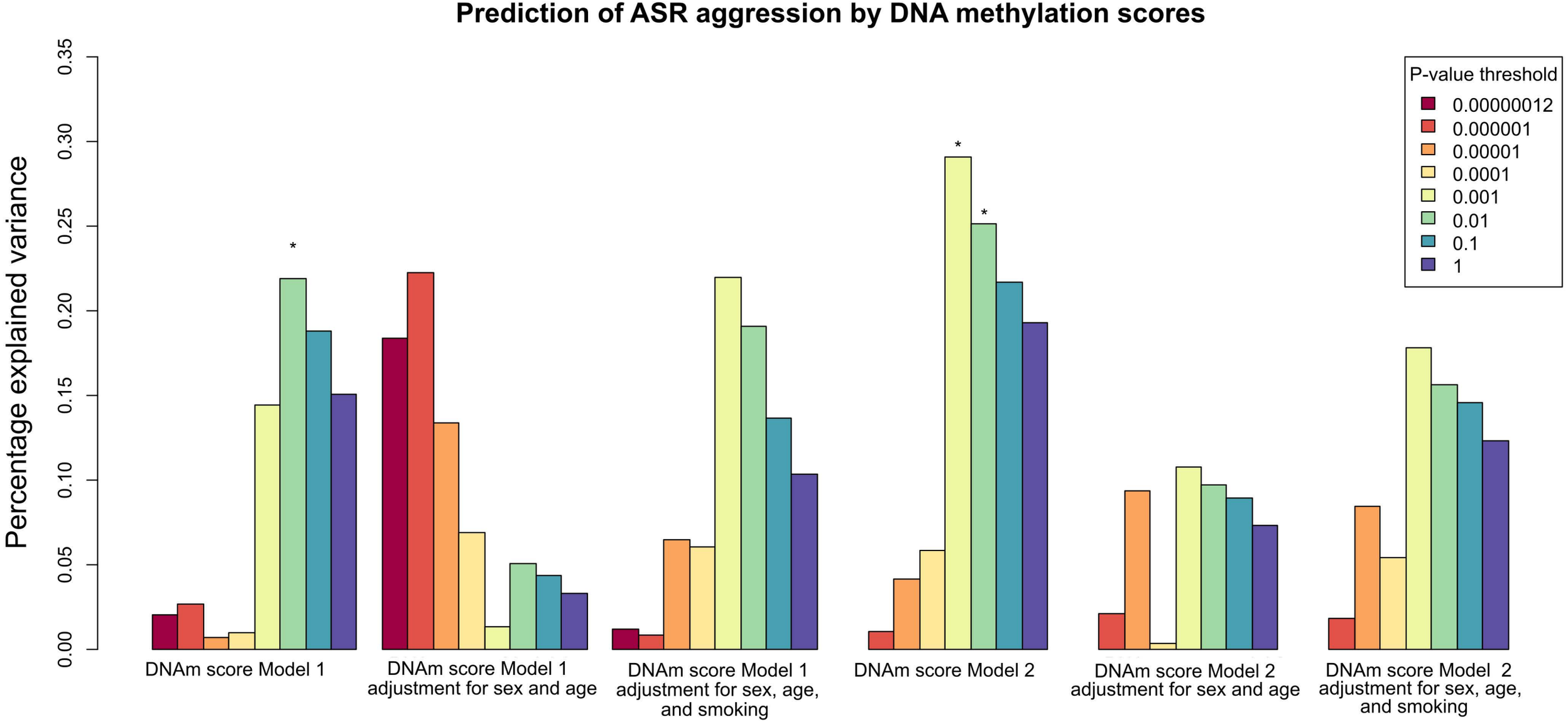

d




c


e

| Trait(s) | -log10(p) | CpGs | % |
|---|---|---|---|
| smoking | >309 | 37 | 0.3 |
| maternal smoking | 197.0 | 22 | 0.7 |
| educational attainment | 172.5 | 21 | 29.2 |
| HIV frailty | 101.0 | 18 | 7.9 |
| smoking cessation | 89.8 | 18 | 2.8 |
| lung function | 88.2 | 14 | 13.3 |
| cognitive function | 74.8 | 15 | 3.8 |
| alcohol consumption | 55.0 | 12 | 0.7 |
| perinatal PCB and PCDF exposure | 48.7 | 7 | 35.0 |
| IgG glycosylation | 46.9 | 6 | 85.7 |


Prediction of ASR aggression by DNA methylation scores
Percentage explained variance
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
P-value threshold
0.00000012
0.000001
0.00001
0.0001
0.001
0.01
0.1
1
*
*
*
DNAm score Model 1
DNAm score Model 1 adjustment for sex and age
DNAm score Model 1 adjustment for sex, age, and smoking
DNAm score Model 2
DNAm score Model 2 adjustment for sex and age
DNAm score Model 2 adjustment for sex, age, and smoking