Integrative multi-omics analysis of childhood aggressive behavior

RUNNING HEAD: Integrative multi-omics analysis of childhood aggressive behavior

Fiona A. Hagenbeek[1,2,*], Jenny van Dongen[1,2,3], René Pool[1,2], Peter J. Roetman[4], Amy C. Harms[5,6], Jouke Jan Hottenga[1], Cornelis Kluft[7], Olivier F. Colins[4,8], Catharina E.M. van Beijsterveldt[1], Vassilios Fanos[9], Erik A. Ehli[10], Thomas Hankemeier[5,6], Robert R. J. M. Vermeiren[4,11], Meike Bartels[1,2], Sébastien Déjean[12], Dorret I. Boomsma[1,2,3]

[1]Department of Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands.

[2]Amsterdam Public Health research institute, Amsterdam, the Netherlands.

[3]Amsterdam Reproduction & Development (AR&D) research institute, Amsterdam, the Netherlands.

[4]LUMC-Curium, Department of Child and Adolescent Psychiatry, Leiden University Medical Center, Leiden, the Netherlands

[5]Division of Analytical Biosciences, Leiden Academic Center for Drug Research, Leiden University, Leiden, the Netherlands

[6]The Netherlands Metabolomics Centre, Leiden, The Netherlands.

[7]Good Biomarker Sciences, Leiden, the Netherlands.

[8]Department Special Needs Education, Ghent University, Ghent, Belgium.

[9]Department of Surgical Sciences, University of Cagliari and Neonatal Intensive Care Unit, Cagliari, Italy.

[10]Avera Institute for Human Genetics, Sioux Falls, South Dakota, USA.

[11]Youz, Parnassia Psychiatric Institute, the Hague, the Netherlands.

[12]Toulouse Mathematics Institute, University of Toulouse, CNRS, Toulouse, France.

* Correspondence: Fiona A. Hagenbeek, Department of Biological Psychology, Vrije Universiteit Amsterdam, Van der Boechorststraat 7-10, 1081 BT Amsterdam, The Netherlands, Telephone: +31 (0)20 59 82974, E-mail: f.a.hagenbeek@vu.nl

N words main text: 7,031 words

N tables: 2 tables

N figures: 3 figures

Abstract

This study introduces and illustrates the potential of an integrated multi-omics approach in investigating the underlying biology of complex traits such as childhood aggressive behavior. In 645 twins (cases=42%), we trained single- and integrative multi-omics models to identify biomarkers for subclinical aggression and investigated the connections among these biomarkers. Our data comprised transmitted and two non-transmitted polygenic scores (PGSs) for 15 traits, 78772 CpGs, and 90 metabolites. The single-omics models selected 31 PGSs, 1614 CpGs, and 90 metabolites, and the multi-omics model comprised 44 PGSs, 746 CpGs, and 90 metabolites. The predictive accuracy for these models in the test (*N*=277, cases=42%) and independent clinical data (*N*=142, cases=45%) ranged from 43% to 57%. We observed strong connections between DNA methylation, amino acids, and parental non-transmitted PGSs for ADHD, Autism Spectrum Disorder, intelligence, smoking initiation, and self-reported health. Aggression-related omics traits link to known and novel risk factors, including inflammation, carcinogens, and smoking.

Keywords: Childhood aggression; multi-omics; polygenic scores; genetic nurturing; DNA methylation; metabolomics.

46 # Introduction

47 Omics studies can lead to an improved understanding of the biological mechanisms contributing
48 to mental health and disorders (Jakovljevic and Jakovljevic 2019). Different omics technologies, e.g.,
49 genomics, epigenomics, transcriptomics, or metabolomics, assess different aspects of the
50 development and progression of complex traits and disorders. The analysis of a single omics layer
51 provides a unique, but likely incomplete, picture of the underlying biology, whereas studies
52 combining multiple omics layers may lead to more comprehensive insights into human biology,
53 because the different omics layers are interrelated and interact (Wörheide et al. 2021). Multi-omics
54 analyses can aid in biomarker discovery, diagnosis, patient classification or subtyping, evaluation of
55 treatment response and uncovering novel insights into disease biology (Pinu et al. 2019;
56 Subramanian et al. 2020).

57 Here, we present an integrative multi-omics analysis of childhood aggressive behavior. Human
58 aggression is a complex and heterogenous behavior encompassing hostile, destructive, or injurious
59 behavior aimed at causing physical or emotional harm to others (Anderson and Bushman 2002;
60 Siever 2008). In several disruptive behavioral disorders, such as conduct and oppositional defiant
61 disorders or intermittent explosive disorder, inappropriate levels of aggressive behavior are
62 observed (Radwan and Coccaro 2020). In humans, high co-occurrence with other social, behavioral,
63 and emotional problems is reported (Bartels et al. 2018; Whipp et al. 2021a). Childhood aggressive
64 behavior puts a burden on children and their parents/caretakers, and is predictive of multiple
65 adverse outcomes later in life, such as antisocial personality disorder (Whipp et al. 2019), criminal
66 convictions (Kassing et al. 2019), lower educational attainment (Vuoksimaa et al. 2020), or negative
67 interpersonal relationships (Fergusson et al. 2005).

68 Large-scale genome-wide association (GWA) studies on aggression have not yet identified
69 significant single nucleotide polymorphisms (SNPs) (Odintsova et al. 2019; Ip et al. 2021). Gene-
70 based analysis for childhood aggression yielded three significantly associated genes (*ST3GAL3*,

71 *PCDH7*, and *IPO13*) (Ip et al. 2021). The polygenic score for childhood aggression does not only
72 significantly explain childhood aggressive behavior at age 7 (Ip et al. 2021), but it also associated
73 with aggression at age 12 to 41 in a Dutch sample and age 38 to 48 in an Australian sample (van der
74 Laan et al. 2021). Small-scale epigenetic studies (sample size range: 41-260) showed that DNA
75 methylation differences in various tissues are associated with aggression and related traits in
76 children and adults (Guillemin et al. 2014; Cecil et al. 2018b, a; Mitjans et al. 2019). The first large-
77 scale epigenome-wide association study (EWAS) meta-analysis across child and adult cohorts (N =
78 15,324) reported 13 significant sites in peripheral blood for broad aggression across the lifespan (van
79 Dongen et al. 2021). Metabolomics studies (sample size range: 77-725) detected plasma and serum
80 metabolites associated with aggression and related traits in adults (Gulsun et al. 2016; Chen et al.
81 2020; Whipp et al. 2021b). A study in 1,347 twins and 183 clinical cases found significant associations
82 of urinary metabolites with childhood aggression (Hagenbeek et al. 2020). These single omics
83 approaches hint at potentially important biological pathways for human aggression, with most of
84 these findings awaiting replication.

85 In this study, we aim to integrate multiple omics layers to construct a multi-omics biomarker
86 panel for childhood aggressive behavior and explore the correlations among the omics traits
87 included in this panel. We collected biological samples in a subproject of ACTION (Aggression in
88 Children: Unraveling gene-environment interplay to inform Treatment and InterventiON strategies):
89 the ACTION Biomarker Study (Boomsma 2015; Bartels et al. 2018; Hagenbeek et al. 2020). The
90 ACTION Biomarker Study comprised a cohort of twins from the Netherlands Twin Register (NTR)
91 (Ligthart et al. 2019) and a clinical cohort of children referred to a youth psychiatry clinic (LUMC-
92 Curium, the Netherlands). The genome-wide SNP and DNA methylation (Illumina EPIC 850K array)
93 data, and the urinary amines and organic acids as measured in these cohorts were previously
94 included in a genome-wide genetic and epigenetic association study and a metabolomics study of
95 (childhood) aggression (Hagenbeek et al. 2020; van Dongen et al. 2021; Ip et al. 2021). We expand
96 on these single-omics studies by integrating these data with a third metabolomics dataset, focusing

97 on urinary steroid hormones. We calculated transmitted and non-transmitted polygenic scores
98 (PGSs) for childhood aggression and a series of genetically correlated traits, such as Attention-Deficit
99 Hyperactivity Disorder (ADHD), smoking, and intelligence. The non-transmitted PGSs assess the
100 indirect effects of genetic variants that were not transmitted from parents to offspring, i.e., genetic
101 nurture, and capture the effects of the environment created by parents beyond the genetic
102 intergeneration transmission (Bates et al. 2018; Kong et al. 2018). Thus, if alleles not transmitted
103 from parents to offspring affect offspring outcomes, this indicates that the offspring's home-
104 environment is influenced by parentental genotypes, and the home-environment in turn affects
105 offspring outcomes (Branje et al. 2020).

106 We employed an analytical design comprising three phases: 1) single-omics analyses; 2) pairwise
107 cross-omics analyses; and 3) multi-omics analyses (**Fig. 1**) (Duruflé et al. 2020). First, we built single-
108 omics biomarker panels in the twin cohort, with 70% of the twin data for model training, 30% of the
109 twin data for model testing, and an independent clinical cohort for follow-up. Second, we examined
110 the pairwise cross-omics connections of those omics traits selected by the single-omics models in
111 the training data. Third, using the same data split for model training and testing, we compared three
112 multi-omics models, with different assumptions on the correlations among the omics traits and
113 describe the multi-omics connections of the selected omics traits.

114 Materials and methods

115 Study population and procedures

116 Participants included 1,494 twins (747 complete pairs) from the Netherlands Twin Register
117 (Ligthart et al. 2019), and 189 children referred to the LUMC-Curium youth psychiatry clinic in the
118 Netherlands that took part in the ACTION Biomarker Study (Aggression in Children: Unraveling gene-
119 environment interplay to inform Treatment and InterventiON strategies) (Boomsma 2015; Bartels et
120 al. 2018; Hagenbeek et al. 2020). Both cohorts collected first-morning urine samples and buccal-cell
121 swabs with standardized protocols (for details, see **Appendix A**). In the twin cohort, we also

122 collected buccal-cell swabs from parents and siblings of the twins. The current study included
123 participants if they had aggression status, complete omics data for all three omics layers, and all
124 relevant covariates (**Table S1**), as the current software does not allow for missing data. Parents
125 provided written informed consent for their children and twin parents provided written informed
126 consent for their own participation. Study approval was obtained from the Central Ethics Committee
127 on Research Involving Human Subjects of the VU University Medical Center, Amsterdam (NTR 25th
128 of May 2007 and ACTION 2013/41 and 2014.252), an Institutional Review Board certified by the U.S.
129 Office of Human Research Protections (IRB number IRB00002991 under Federal-wide Assurance-
130 FWA00017598; IRB/institute codes), and the Medical Ethical Committee of Leiden University
131 Medical Center (B17.031, B17.032 and B17.040).

132 Aggressive behavior

133 Parents and teachers rated aggressive behavior on the Aggressive Behavior syndrome scale of the
134 Achenbach System of Empirically Based Assessment (ASEBA) Child Behavior Checklist (CBCL) or
135 Teacher Report Form (TRF) (Achenbach et al. 2017). We have described selection and definition of
136 aggression cases and controls previously (Hagenbeek et al. 2020). In brief, we selected twin pairs
137 based on concordance (case-case or control-control) and discordance (case-control pairs) for
138 mother- (93%) or teacher-rated (7%) aggressive behavior at ages 3, 7, and/or 10 years. We matched
139 concordant control pairs on postal code to the case-case and case-control pairs, and collection of
140 biological samples within regions in the Netherlands was around the same time. Cases were defined
141 by mother- or teacher-rated sex- and age-specific T-scores of 65 or higher (subclinical levels), or on
142 mother-rated age-specific thresholds on the item scores at age 3 (≥ 13), age 7 (≥ 5), or age 10 (≥ 4)
143 (N = 388, **Table S1**). We denoted individuals with scores below these thresholds as controls (N = 534,
144 **Table S1**). In the clinical cohort, children with a parent-rated sex-specific T-score of ≥ 70 (clinical
145 levels) were classified as cases (N = 64), and children with T-scores of < 65 were classified as low
146 scoring controls (N = 78, **Table S1**). We excluded children with T-scores in the subclinical range (T-
147 scores of ≥ 65 and < 70) from this study (N = 35).

Omics measurements

Detailed information on omics measurements is included in **Appendix B**. Genotyping was performed on Affymetrix Axiom or Illumina GSA arrays (Ehli et al. 2017; Beck et al. 2019), and genome-wide SNP data were available for 3,334 participants, including 1,702 parents and siblings of twins (AXIOM = 909, GSA = 2,425). Transmitted and two non-transmitted polygenetic scores (PGS) were calculated for childhood aggression and 14 other traits that showed a significant ($p < 0.02$) genetic correlation of ≤ -0.40 or ≥ 0.40 with childhood aggression (Ip et al. 2021) (**Table 1**). Thus, in total we calculated 45 PGSs: a transmitted, non-transmitted maternal, and non-transmitted paternal PGS for each trait. The effects of sex, age at biological sample collection, genotyping platform, and the first 10 genetic principal components (PCs) were regressed on the standardized PGSs and we included residuals in the analyses.

Genome-wide DNA methylation was measured on the Infinium MethylationEPIC BeadChip Kit (Illumina, San Diego, CA, USA (Moran et al. 2016)). Quality Control (QC) and normalization were carried out with pipelines developed by the Biobank-based Integrative Omics Study (BIOS) consortium (Sinke et al. 2019). From the 787,711 autosomal methylation probes that survived QC, the top 10% most variable probes were included in the analyses (**Data S1**). Residual methylation levels were obtained by regressing the effects of sex, age, percentages of epithelial and natural killer cells, EPIC array row, and bisulfite sample plate, from the methylation β-values.

Urinary metabolomics data were generated on: 1) a liquid chromatography mass spectrometry (LC-MS) platform targeting amines; 2) a LC-MS platform targeting steroid hormones; and 3) a gas chromatography (GC) MS platform targeting organic acids. We excluded metabolites with a relative standard deviation of the QC samples larger than 15% and retain 60 amines, 10 steroids, and 20 organic acids in our analyses. After QC, we normalized metabolite levels to the sample-median and inverse normal rank transformed. We analyzed residuals obtained by regressing the effects of sex and age on the normalized and transformed urinary metabolites.

173 ## Statistical analyses

174 To define multi-omics biomarker panels capable of discriminating between cases and controls
175 and explore the connections among the omics traits included in these panels, we employed an
176 analytical design comprising three phases: 1) single-omics analyses; 2) pairwise cross-omics analyses;
177 and 3) multi-omics analyses (**Fig. 1**). To avoid overfitting of the single- and multi-omics models, we
178 randomly split the twin sample at the twin pair level into two subsets: 70% of the data for model
179 training (training data), and 30% of the data for model testing (test data; **Table 2**). The clinical
180 validity of the final single- and multi-omics models was evaluated in the clinical cohort (**Table 2**). We
181 carried analyses out in the mixOmics R package (version 6.12.1) implemented in the R programming
182 language (version 4.0.2) (R Core Team; Rohart et al. 2017).

183 ### Phase 1: Single-omics analyses

184 #### *Univariate polygenic score analyses*

185 In the training data, we first associated the transmitted- and non-transmitted PGSs with
186 childhood aggression through generalized estimating equation (GEE) models. GEE models tested the
187 association of each transmitted and non-transmitted PGS separately on the continuous mother-
188 rated sum scores of the ASEBA CBCL Aggressive Behavior syndrome scale as assessed at the time of
189 biological sample collection. All models included sex, age, genotype array, and the first 10 genetic
190 principal components as covariates. We corrected for the correlation structure within families by
191 using the "exchangeable" correlation structure, obtaining robust variance estimators (Rogers and
192 Stoner 2016). A False Discovery Rate (FDR) of 5% for 45 PGSs was used to correct for multiple testing
193 (p.adjust function in R), setting the significance threshold to $q \leq 0.05$, i.e., 5% of the significant
194 results will be false positives (Benjamini and Hochberg 1995).

195 #### *Multivariate single-omics models*

196 To get a first insight into the dimensionality of the metabolomics data (90 variables), DNA
197 methylation data (78,772 variables), and the PGSs (45 variables), we ran Principal Component

198 Analysis (PCA) within each omics block in the training data (**Table S2**). To assess the ability of each of
199 the three omics layers to correctly classify aggression status, we applied Partial Least Square
200 Discriminant Analysis (PLS-DA) in the training data. PLS-DA involves iteratively constructing
201 successive latent components, where each component is a linear combination of the included omics
202 variables (Rohart et al. 2017). For each component, PLS-DA aims to maximize the covariance
203 between the residual X matrix, containing the omics data, and the Y matrix, containing the sample
204 classification (i.e., case-control status coded as a dummy variable). The mixOmics software requires
205 a user-defined maximum number of components in PLS-DA models. We chose this maximum based
206 on the number of PCs as determined by the elbow method in the PCA (**Fig. S1**; **Table S2**). To find the
207 optimal number of components to keep in each PLS-DA model, we employed 10-fold cross-validation
208 (CV) with 100 repeats (*perf* function; **Table S3**; **Fig. S2**).

209 Next, we applied sparse PLS-DA (sPLS-DA) to reduce the number of variables in each omics block
210 contributing to each component. sPLS-DA includes Least Absolute Shrinkage and Selection Operator
211 (LASSO) penalization (e.g., *L1* penalization). LASSO shrinks the coefficients of less importance, often
212 highly correlated, variables to zero, removing these variables from the model (Tibshirani 1996). Thus,
213 for each component, sPLS-DA finds the maximum covariance between the residual X matrix
214 containing a subset of the omics data with non-zero coefficients (variable selection) and the Y matrix
215 (Lê Cao et al. 2011). We assessed the variable selection via 10-fold CV with 100 repeats (*tune*
216 function), keeping at least two components in the final model (**Table S3**; **Fig. S3**). CV again obtained
217 the performance of the final sPLS-DA model using the *perf* function (10 folds, 100 repeats; **Table S3**;
218 **Fig. S4**).

219 The ability of the final single-omics models to predict out-of-sample case-control status was
220 evaluated in the test and clinical datasets (*predict* function). For each new observation in the test
221 and clinical datasets, this function calculates the predicted class (case/control) by estimating their
222 predicted dummy variable (of the case-control status) using the maximum, Mahalanobis, or

223 Centroids (Euclidian) distance (see Rohart et al. 2017). When using the maximum distance, the
224 predicted class of a new observation is the class for which we observed the largest predicted dummy
225 value. Both the Mahalanobis and centroids distances are centroid-based distances that predict the
226 class of a new observation, so that the distance between its centroid and predicted scores is
227 minimal. To predict out-of-sample case-control status, we used the best performing prediction
228 distance, as was determined during model training (see **Table S3**).

229 The misclassification rates of the models, that combine the number of cases classified as controls
230 (false negative rate) and the number of controls classified as cases (false positive rate), was used to
231 evaluate how well the final models predicted case-control status. We employed a balanced
232 misclassification rate, the balanced error rate (BER), that corrects for imbalances in the number of
233 cases and controls. We used a confusion matrix, comparing the true cases and controls with the
234 predicted cases and controls, to calculate the sensitivity (number of cases correctly classified as
235 cases [true positive rate]), specificity (number of controls correctly classified as controls [true
236 negative rate]), and accuracy (overall correct classification). As an alternative, Receiver Operating
237 Characteristic (ROC) analysis assessed the Area Under the Curve (AUC) in both the test and clinical
238 data. We obtained the ROC curve per component.

239 Phase 2: Pairwise cross-omics analyses

240 To highlight pairwise cross-omics relationships (i.e., DNA methylation-metabolomics, PGSs-
241 metabolomics, and PGSs-DNA methylation), Partial Least Squares (PLS) regression models were
242 constructed in canonical mode in the training data. Similar to canonical correlation analysis,
243 canonical PLS regression aims to find linear combinations of the variables (canonical variates) to
244 reduce the dimensionality of the data while maximizing the covariance between the variates (Rohart
245 et al. 2017). CV has not yet been implemented for PLS in canonical mode. Consequently, we kept the
246 smallest number of components as kept in either of the respective single-omics models (**Table S3**).

247 Therefore, we included 3 components for the DNA methylation-metabolomics model, and 2
248 components for the PGS-metabolomics and PGS-DNA methylation models.

249 The PLS models were run with two different sets omics variables, 1) a PLS model that included all
250 45 PGSs, 78,772 CpGs, and 90 metabolites (model 1), and 2) a PLS model that only included the 36
251 PGSs, 1,614 CpGs, and 90 metabolites that were selected by the single-omics sPLS-DA models
252 (model 2; **Data S2**). Model 1 provides insight into the connections among all omics variables, while
253 model 2 provides insight into the connections among those omics variables that best contribute to
254 aggression case-control classification in the single-omics sPLS-DA models. In the remainder of the
255 text we will refer to these connections as 'correlations', but note that these are not Pearson
256 correlations between among omics variables. These connections are the correlations between the
257 projected variables onto the space spanned by the components as retained in the analysis, i.e., the
258 correlation among the PLS variates (González et al. 2012). On the omics variables included in model
259 2, we performed hierarchical clustering with the Ward linkage algorithm on Euclidean distances of
260 the PLS variates and used the 'dendextend' R-package (Galili 2015) to extract the two largest clusters
261 for both of the omics blocks included in the PLS models.

262 Phase 3: Multi-omics analyses

263 The multi-omics analysis was conducted through Data Integration Analysis for Biomarker
264 discovery using Latent cOmponents (DIABLO) in the training data. DIABLO extends PLS-DA to multi-
265 block PLS-DA (MB-PLS-DA), that aims at identifying correlated variables from multiple omics blocks
266 that maximize the sample classification (Singh et al. 2019). The method requires a user-defined
267 'design matrix', that specifies the expected correlations among the omics blocks. The symmetric
268 design matrix has the number of rows and columns equal to the number of omics blocks (i.e., 3), and
269 contains values between 0 and 1. A 'full' design matrix denotes strong positive correlations among
270 the omics blocks and sets the values among omics blocks close to or equal to one. A 'null' design
271 matrix denotes weak or no correlations among omics blocks by setting values close to or equal to

272 zero. The full design matrix optimizes correlations among the omics blocks, while the null design
273 matrix optimizes the discrimination between samples (Rohart et al. 2017; Singh et al. 2019; Duruflé
274 et al. 2020). We can also specify a design matrix with the empirical correlations among the omics
275 blocks.

276 We compared a multi-omics model with an empirical design matrix (based on the correlations
277 obtained from the model 1 pairwise cross-omics PLS models; **Table S4**) to models with a null or full
278 design matrix. Based on the results of the single-omics sPLS-DA models (**Table S5**), we chose the
279 maximum number of components to include in the MB-PLS-DA models. We determined the optimal
280 number of components to keep in the MB-PLS-DA model with 10-fold CV and 100 repeats (*perf*
281 function; **Fig. S5**). We assessed the variable selection per component per omics block via 5-fold CV
282 with 50 repeats (*tune* function; **Table S5**; **Fig. S6**). Performance of the final MB-sPLS-DA model was
283 assessed with 5-fold CV (50 repeats; **Table S5**; **Fig. S7**).

284 The ability of the final multi-omics models to predict out-of-sample case-control status was
285 evaluated in the test and independent clinical data (*predict* function), based on the best performing
286 prediction distance as was determined during model training (see **Table S5**). The final multi-omics
287 models were evaluated by their balanced error rates (BER), and the sensitivity, specificity, and
288 accuracy of the models were calculated from their confusion matrices. In the multi-omics models,
289 we calculated the ROC curves per component for each omics block.

290 Biological characterization

291 To facilitate biological interpretation, we describe the connections of the PGSs, CpGs, and
292 metabolites, that were selected by the single-omics sPLS-DA models and those selected by the multi-
293 omics MB-sPLS-DA models. For the multi-omics MB-sPLS-DA models we also identified correlation
294 patterns that included high absolute correlations ($|r| \geq 0.60$) between omics traits of at least two
295 omics blocks. As for the pairwise cross-omics models, these correlations comprise the correlations
296 among the PLS variates. To construct the correlational patterns we denoted all instances where

297 multiple omics variables were connected through bi- or trivariate correlations as a pattern. An
298 example of a pattern comprising one variable of each omics block: high correlation of a PGS with a
299 CpG and a high correlation of this CpG with a metabolites.

300 To test for enrichment of methylation sites previously associated with other traits, we performed
301 trait enrichment analysis for all traits (619) in the EWAS atlas on the 18th of June 2021 (Xiong et al.
302 2022). CpGs served as input for the trait enrichment analysis if 20 or more unique CpGs were
303 selected into the single-omics sPLS-DA model, the multi-omics MB-sPLS-DA models, or included in a
304 multi-omics MB-sPLS-DA high correlation pattern. When fewer than 20 CpGs were selected, we
305 manually retrieved the trait associations with the CpGs from the EWAS atlas. Similarly, we
306 performed trait enrichment analysis or manual retrieval of the CpGs included in the clusters as
307 identified for the pairwise cross-omics analyses for all traits (618) in the EWAS atlas on the 1st of July
308 2021.

309 Results

310 Polygenic prediction

311 Transmitted and non-transmitted PGSs for childhood aggression and 13 other traits were
312 individually not significantly associated with aggressive behavior after multiple testing correction
313 (**Data S3**). The transmitted PGS for ADHD were significantly associated with aggressive behavior (β =
314 1.16, *SE* = 0.26, q = 0.0003), while the non-transmitted PGSs were not (mother: β = -0.22, *SE* = 0.24, q
315 = 0.83; father: β = 0.02, *SE* = 0.23, q = 0.98), showing that genetic liability for ADHD associates with
316 increased levels of aggressive behavior.

317 Single-omics models for childhood aggression

318 We built cross-validated single-omics biomarker panels for childhood aggressive behavior based
319 on sPLS-DA models including PGSs, DNA methylation, or metabolomics data. The final models
320 comprised 11 transmitted and 25 non-transmitted PGSs, 1,614 CpGs, and all 90 metabolites (**Table**

321 **S3**; **Data S2; Fig. S8**). The PGSs selected in the sPLS-DA model comprised the transmitted PGSs for
322 aggression, ADHD, age at first birth, age at smoking initiation, number of cigarettes per day,
323 educational attainment (EA), insomnia, loneliness, Major Depressive Disorder (MDD), smoking
324 initiation, and wellbeing. The non-transmitted maternal PGSs were selected for aggression, ADHD,
325 age at first birth, age at smoking initiation, Autism Spectrum Disorder (ASD), childhood IQ, number
326 of cigarettes per day, EA, insomnia, intelligence, loneliness, MDD, self-reported health, smoking
327 initiation, and wellbeing, and the non-transmitted paternal PGSs were selected for aggression,
328 ADHD, ASD, childhood IQ, number of cigarettes per day, intelligence, MDD, self-reported health,
329 smoking initiation, and wellbeing (**Data S2**). Trait enrichment analyses for all 1,614 selected CpGs in
330 the sPLS-DA model showed the strongest enrichment for glucocorticoid exposure (i.e.,
331 administration of corticosteroid medication (Braun et al. 2019); OR = 18.02, $p = 5.34 \times 10^{-158}$), and
332 household socioeconomic status in childhood (OR = 9.88, $p = 1.18 \times 10^{-13}$; **Table S6**). All three single-
333 omics models showed low classification accuracy in the test and clinical data (**Table S7**; **Fig. S9- S10**).

334 Pairwise cross-omics models

335 DNA methylation-metabolomics

336 In the pairwise DNA methylation-metabolomics model, we used hierarchical clustering and found
337 two clusters of CpGs and of metabolites (**Fig. 2a**; **Data S4**). The DNA methylation cluster 1 contains
338 1,151 (71.3%) of the CpGs selected by the sPLS-DA models for childhood aggression, and showed the
339 strongest trait enrichments for household socioeconomic status in childhood (OR = 12.48, $p =$
340 4.20×10^{-14}; **Table S8**). The 463 (28.7%) CpGs included in cluster 2 showed the strongest trait
341 enrichment for glucocorticoid exposure (OR = 51.25, $p \leq 1.00 \times 10^{-308}$; **Table S8**). Metabolite cluster 1
342 contains 69 (76.7%) of the metabolites, including 55 amines, 7 organic acids and 7 steroids, while
343 metabolite cluster 2 contains 21 metabolites (23.3%), including 5 amines, 13 organic acids, and 3
344 steroids. The average correlation between the DNA methylation and metabolomics blocks, i.e., the
345 correlation among all PLS variates of all components simultaneously, was 0.18 ($q = 6.07 \times 10^{-15}$). We

346 observed the highest absolute correlations between metabolites of cluster 1 and CpGs included in
347 cluster 2 (**Fig. 2a**; **Data S4-S5**). Specifically, the amines 3-methoxytyramine (*r* M = -0.21, r range: -
348 0.23 to 0.20), asymmetric dimethylarginine (ADMA, *r* M = -0.20, *r* range: -0.21 to -0.20), L-glutamic
349 acid (*r* = -0.20), L-phenylalanine (*r* M = -0.20, *r* range: -0.21 to -0.20), O-acetyl-L-serine (*r* = -0.20), and
350 symmetric dimethylarginine (SDMA, *r* M = -0.21, *r* range: -0.22 to -0.20) show negative correlations
351 with these CpGs, indicating that increased levels of these urinary metabolites associated with
352 hypomethylation at cluster 2 CpG sites.

353 ## PGSs-metabolomics

354 Hierarchical clustering of the pairwise PGSs-metabolomics model identified two clusters of PGSs
355 and two clusters of metabolites (**Fig. 2b**; **Table S9**). The PGS cluster 1 contains 23 of the PGSs (63.9%)
356 selected by the sPLS-DA model for childhood aggression. This cluster comprises 8 transmitted PGSs,
357 9 non-transmitted maternal PGSs, and 6 non-transmitted paternal PGSs. The 13 PGSs (36.1%)
358 included in cluster 2 comprised 3 transmitted PGSs, 6 non-transmitted maternal PGSs, and 4 non-
359 transmitted paternal PGSs. The metabolite cluster 1 contains 46 metabolites (51.1%), all of which
360 were amines (76.7% of all amines). The remaining 14 amines, as well as all steroids and organic
361 acids, are included in metabolite cluster 2 (*N* = 44, 48.9%). The average correlation between the
362 PGSs and metabolomics block, i.e., the correlation among all PLS variates of all components
363 simultaneously, was 0.28 ($q = 9.57 \times 10^{-24}$). We observed the highest positive correlations between
364 the non-transmitted paternal ADHD (ADHD_NTf) and smoking initiation PGSs
365 (smokinginitiation_NTf) with both essential and non-essential amino acids (ADHD_NTf *r* M = 0.21, *r*
366 range: 0.21-0.22, smokinginitiation_NTf *r* M = 0.22, *r* range: 0.20-0.26; **Fig. 2b**; **Data S6**; **Table S9**).
367 Similarly, the most negative correlations were observed between amino acids and the non-
368 transmitted paternal PGSs for intelligence (intelligence_NTf) and self-reported health
369 (selfreportedhealth_NTf, intelligence_NTf *r* M = -0.21, *r* range: -0.22 to -0.21, selfreportedhealth_NTf
370 *r* M = -0.21, *r* range: -0.22 to -0.20). This shows that characteristics tagged by ADHD, smoking

371 initiation, intelligence, and self-reported health correlated to paternal construction of environments
372 influencing offspring urinary amino acid levels.

373 PGSs-DNA methylation

374 For the pairwise PGSs-DNA methylation model, we again applied hierarchical clustering to find
375 two clusters for the PGSs and DNA methylation data (**Fig. 2c**; **Data S7**). PGS cluster 1 contains 22 of
376 the PGSs (61.1%) selected by the sPLS-DA model for childhood aggression. This cluster contained 6
377 transmitted PGSs, 9 non-transmitted maternal PGSs, and 7 non-transmitted paternal PGSs. The 14
378 PGSs (38.9%) included in cluster 2 comprised 5 transmitted PGSs, 6 non-transmitted maternal PGSs,
379 and 3 non-transmitted paternal PGSs. The DNA methylation cluster 1 contained 1,142 (70.8%) of the
380 CpGs selected by the sPLS-DA models for childhood aggression, and showed the strongest trait
381 enrichments for household socioeconomic status in childhood (OR = 12.58, $p = 3.68 \times 10^{-14}$; **Table**
382 **S10**). The 472 (29.2%) CpGs included in cluster 2 showed the strongest trait enrichment for
383 glucocorticoid exposure (OR = 50.00, $p \leq 1.00 \times 10^{-308}$; **Table S10**). The average correlation between
384 the PGSs and DNA methylation block, i.e., the correlation among all PLS variates of all components
385 simultaneously, was 0.28 ($q = 9.57 \times 10^{-24}$). The most negative correlations were observed between
386 the non-transmitted paternal self-reported health PGS with CpGs of cluster 2 (*r* M = -0.18, *r* range: -
387 0.19 to -0.15), and the highest positive correlations between the cluster 2 CpGs and the non-
388 transmitted paternal ASD PGS (*r* M = 0.151, *r* range: 0.150-0.154; **Fig. 2c**; **Data S8**). This indicates that
389 characteristics tagged by ASD and self-reported health correlated to paternal construction of
390 environments influencing offspring buccal DNA methylation levels.

391 Multi-omics model for childhood aggression

392 We built multi-omics panels for childhood aggressive behavior based on multi-block sPLS-DA
393 (MB-sPLS-DA) models, including PGSs, DNA methylation, and metabolomics data. Here, we report
394 the multi-omics model with an empirical design matrix and results for the null and full design
395 matrices can be found in **Appendices C** and **D**, respectively. After cross-validation, the optimal 5-

396 component model included 14 transmitted and 30 non-transmitted PGSs, 746 CpGs, and all 90
397 metabolites (**Table S5**; **Data S9**; **Fig. S11**). The multi-omics model selected all 30 non-transmitted
398 PGS, and except for the transmitted PGS for childhood IQ, also all the transmitted PGSs (**Data S9**).
399 Out of the 746 CpGs selected by the multi-omics model 204 (27.3%) were also selected by the single-
400 omics DNA methylation model (**Data S2; Data S9**). Trait enrichment analyses for all selected CpGs in
401 the MB-sPLS-DA model showed the strongest enrichment for gender (OR = 3.90, $p = 3.97 \times 10^{-31}$), and
402 breast cancer risk (OR = 342.60, $p = 1.44 \times 10^{-21}$; **Table S11**). Multi-omics prediction of aggression
403 case-control status in the test data showed an improvement in the prediction (BER = 0.47-0.52;
404 **Table S12**; **Fig. S12**) as compared to single-omics models including only the PGSs or metabolomics
405 data, but not as compared to the model including only the DNA methylation data (**Table S7**). In
406 contrast, in the clinical cohort, the average classification accuracy was poorer in the multi-omics
407 model (BER = 0.53-0.57; **Table S12**) than for the single-omics models (**Table S7**; **Fig. S13**).

408 The average correlations between each omics block in the multi-omics model, i.e., the
409 correlation among all PLS variates of all components simultaneously, were $r = 0.19$ ($q = 2.13 \times 10^{-27}$)
410 for PGSs-DNA methylation, $r = 0.13$ ($q = 1.78 \times 10^{-12}$) for PGSs-metabolomics, and $r = 0.15$ ($q =$
411 3.76×10^{-16}) for DNA methylation-metabolomics. We observed high absolute correlations ($|r| \geq 0.60$)
412 between 2 selected PGSs, 14 CpGs, and 13 metabolite, that we summarize in five sets of
413 correlational patterns (**Fig. 3**; **Data S10; Table S13**). Correlation pattern 1 comprises high negative
414 correlations of citric and fumaric acid with cg12886033 (chr1:65449013), cg14508705
415 (chr1:172360182), and cg11710553 (chr4:105892960), and high negative correlations of fumaric acid
416 with cg21432062 (chr3:4908643), cg22848658 (chr6:135354586), and cg15841349
417 (chr12:129348564). The second correlational pattern is characterized by high negative correlations
418 between isocitrate with cg05056638 (chr8:24800824), cg08415582 (chr8:57030523), cg11206167
419 (chr5:42924367), and cg20704654 (chr20:30072118). Correlation pattern 3 contains the high
420 positive correlation of homocysteine with cg13784456 (chr10:132970405) and cg06144718
421 (chr10:133048392). The fourth correlational pattern includes high positive correlations between

422 cg03469862 (chr11:68924853) with the transmitted PGS for ADHD, 3-methoxytyrosin, L-isoleucine,
423 L-leucine, L-phenylalanine, L-tryptophan, L-valine, L-glutamine, L-tyrosine, and L-serine. The final
424 correlation pattern comprises a high negative correlation between the non-transmitted by father
425 PGS for EA and cg09674340 (chr1:202509286).

426 Discussion

427 This study comprised an integrative multi-omics analysis of childhood aggressive behavior to
428 identify a multi-omics biomarker panel and to investigate the correlations among the omics blocks
429 included in this panel. In our training data comprising 645 twins (cases = 42.0%, controls = 58.0%) we
430 applied multivariate statistical methods to analyze and integrate transmitted and paternal and
431 maternal non-transmitted PGSs for childhood aggression and for 14 traits genetically correlated with
432 aggression (45 PGSs total), 78,772 CpGs, and 90 metabolites (**Fig. 1**). We build single-omics
433 biomarker panels for each of the omics blocks, that selected 31 PGSs, 1,614 CpGs, and 90
434 metabolites to discriminate between aggression cases and controls. The markers selected in single-
435 omics models had poor predictive performance in our test (N = 277, cases = 42.2%, controls = 57.8%)
436 and clinical datasets (N = 142, cases = 45.1%, controls = 54.9%). We explored the pairwise
437 correlations of the omics variables selected by the single-omics models and found that indirect
438 genetic effects for ADHD, ASD, intelligence, smoking initiation, and self-reported health, connected
439 most strongly with buccal DNA methylation and urinary amino acid levels in children and that higher
440 amino acid levels associate with hypomethylation in a cluster of CpG sites.

441 The multi-omics panel selected 44 PGSs, 746 CpGs, and 90 metabolites, which also had poor
442 predictive performance in our test and clinical datasets. We described five sets of correlational
443 patterns with high absolute correlations ($|r| \geq 0.60$) of aggression-related omics variables selected
444 by the multi-omics model. Correlation patterns 1 and 3 likely capture the DNA methylation process
445 itself, as homocysteine is involved in the methionine cycle that transfers the methionine methyl
446 group to DNA methyltransferase (Selhub 1999), and tricarboxylic acid (TCA) cycle metabolites, such

447 as citric and fumaric acid, play a role in histone acetylation, and histone and DNA demethylation
448 (Martínez-Reyes and Chandel 2020). The CpGs and their connected metabolites included in patterns
449 1, 2, and 4 were previously associated with numerous traits, including ageing, inflammatory bowel
450 disease (IBD), prostate cancer, and Myelodysplastic Syndrome (Ooi et al. 2011; Schicho et al. 2012;
451 Dawiskiba et al. 2014; Rist et al. 2017; Lima et al. 2021; Yuan et al. 2021; Xiong et al. 2022). Pattern 5
452 links a smoking and low birth weight-associated CpG to indirect genetic effects for EA. The negative
453 association of a smoking-associated CpG with indirect genetic effects for EA may indicate that
454 rearing environment influences smoking-related DNA methylation. Pattern 5 may be regarded as
455 being in line with a previous study that reported significant negative correlations of indirect parental
456 EA effect with offspring smoking initiation, number of cigarettes per day, and smoking cessation (Wu
457 et al. 2021), and the well-established phenotypic association of low paternal educational attainment
458 with low offspring birth weight (e.g., Meng and Groth 2018). Overall, the multi-omics correlational
459 patterns associated with a range of traits that link aggression-related omics variables to biological
460 processes related to inflammation, carcinogens, ageing, sex differentiation, intelligence, and
461 smoking.

462 We explored how sensitive our results are to different specifications of the design matrices and
463 found that these resulted in similar predictive abilities, but selected different numbers of only
464 partially overlapping PGSs and CpGs (**Appendices C** and **D**). In line with expectations, the predictive
465 ability of the null design matrix was slightly better than for the empirical design matrix, reflecting
466 that such a model focuses on selecting discriminatory variables (Singh et al. 2019; Duruflé et al.
467 2020). We expected that the model with the full design model sacrificed predictive accuracy to
468 select discriminatory variables that are also highly correlated, but this model had the lowest
469 classification error rate as compared to the models with an empirical or null design matrix in the test
470 and clinical data. Multi-omics models with differently specified design matrices differed not only in
471 how many and which PGSs and CpGs were selected, but each multi-omics model provided unique
472 insight into the correlations among the omics variables selected for their association with childhood

473 aggression. In the current study, we relied on the cross-validation results to select the number of
474 variables and components in the model. However, to aid in biological interpretation, other choices
475 can be made.

476 The single- and multi-omics models selected nearly all the PGSs, with only the transmitted PGSs
477 for ASD, intelligence, childhood IQ, and self-reported health, and the non-transmitted paternal PGSs
478 for age at first birth, EA, insomnia, and loneliness not selected to the single-omics model, and the
479 transmitted PGS for childhood IQ not selected to the multi-omics model (empirical design matrix).
480 The high selection of transmitted PGSs to the biomarker panels likely reflect the genetic correlations
481 of these traits with childhood aggression on which basis we included them in the current study (Ip et
482 al. 2021). The parental non-transmitted PGSs in the genomics block capture the effect of parental
483 genotypes on their offspring's rearing environment, without confounding by genetic transmission.
484 The non-transmitted PGSs for various traits, including childhood aggression, were consistently
485 retained in the single- and multi-omics models, which is in line with research that shows associations
486 of parenting styles with childhood aggression (Masud et al. 2019).

487 The CpGs in the single- or multi-omics models did not overlap with the top differentially
488 methylated CpGs for physical aggression as observed in buccal-cells (Cecil et al. 2018b), and were
489 not significantly associated with aggression in a recent blood-based EWAS meta-analysis (van
490 Dongen et al. 2021). That our models selected no CpGs overlapping with these studies was
491 unsurprising, since it was previously shown that the top CpGs from these studies were not
492 associated with aggressive behavior in the buccal DNA methylation data in the NTR and LUMC-
493 Curium cohorts (van Dongen et al. 2021). Trait enrichment of the CpGs selected by the single- and
494 multi-omics models reported enrichment of known aggression risk factors, such as socioeconomic
495 status (Miller and Tolan 2019; Bellair et al. 2019; Hendriks et al. 2020), childhood malnutrition (Liu
496 2004; Vaughn et al. 2016), and pre- and perinatal risk factors (Van Adrichem et al. 2020). Moreover,
497 we observed trait enrichment for several syndromes, e.g., Down syndrome, that are characterized

498 by intellectual disability, developmental delay, and sometimes by aggressive behaviors among
499 affected individuals. Several of these syndromes, including Klinefelter and Claes-Jensen syndrome,
500 are linked to genetic abnormalities of the sex chromosomes, and CpGs in the single- and mult-omics
501 models are enriched for steroid hormone-related traits, such as glucocorticoid exposure. Together,
502 these results suggest that the CpGs in the single- and multi-omics models reflect the well-established
503 gender differences in aggressive beavior. In addition to the enrichment of glucocorticoid exposure in
504 CpGs selected by the single-omics model, we observed a high correlation between cortisol and
505 cg05153029 (chr20:19769815, r = 0.61), that associates with glucocorticoid exposure in the EWAS
506 atlas (full design matrix MB-sPLS-DA model). Cortisol and other hypothalamus-pituitary-adrenal
507 (HPA)-axis molecules have been implicated in aggressive behavior, suggesting a role of the stress
508 response system in aggressive behavior (Hagenbeek et al. 2016). Epigenetic programming of the
509 HPA-axis is influenced by pre- and perinatal factors, such as maternal behavior as observed in rats
510 (Weaver et al. 2004), and maternal stress and early life adversity (Mulligan et al. 2012; Hompes et al.
511 2013; Jiang et al. 2019), and lower average global DNA methylation levels were reported in patients
512 with Cushing's Syndrome (CS) in remission, a model for long-standing excessive glucocorticoid
513 (cortisol) exposure (Glad et al. 2017). Thus, epigenetic mechanisms may mediate the association
514 between cortisol and childhood aggression, similarly as the DNA methylation mediates the
515 association between cortisol stress reactivity and childhood trauma (Wrigglesworth et al. 2019).

516 The single- and multi-omics models selected all 90 LC- and GC-MS metabolites for inclusion,
517 which might be explained by the low expression differences between aggression cases and controls
518 of the urinary amines and organic acids in this sample (Hagenbeek et al. 2020). Another plausible
519 explanation regards the fact that we generated the metabolomics data on three targeted platforms,
520 chosen because they cover relevant metabolites involved in neurotransmitter, inflammation, and
521 steroid hormone pathways that associate with aggression (Hagenbeek et al. 2016). To date, no
522 metabolomics platform can capture the entire metabolome. In addition, within a platform, technical
523 challenges may cause compounds becoming undetectable or not quantifiable. This was the case for

524 the steroid platform being less successful in quantifying (conjugated) sex hormones. Coverage could
525 also be extended to include other metabolite classes, preferably in a hypothesis-free manner by
526 employing non-targeted metabolomics platforms. The five sets of omics variables with high cross-
527 omics correlations in the multi-omics model contained 13 metabolites. These metabolites comprised
528 8 amino acids, 3 metabolites involved in the TCA cycle, the dopaminergic trace amine 3-
529 methoxytyrosine, and homocysteine, which is involved in cysteine and methionine metabolism. This
530 is in line with previous studies reporting association of metabolites with aggressive behavior (Gulsun
531 et al. 2016; Hagenbeek et al. 2016, 2020; Chen et al. 2020).

532 This is the first multi-omics study that includes DNA methylation profiles from buccal and
533 metabolomics in urine. Most of the earlier large-scale omics studies were conducted in blood
534 samples. By obtaining omics measurements in easily accessible peripheral tissues (urine and buccal-
535 cells) we could obtain multi-omics data in a substantially larger (~10-fold) sample compared to most
536 previous multi-omics studies for psychiatric traits, such as depression and suicide risk (Bhak et al.
537 2019) or post-traumatic stress disorder (Dean et al. 2019), that relied on small training samples
538 (range: 126-165). In the current study, we corrected the omics data for sex, age, and technical
539 covariates, which can be considered as a minimal set of covariates. Future studies may consider
540 including other confounding factors, such as parental smoking, dietary factors, or body mass index.
541 In evaluating the validity of the PGS and multi-omics models in the clinical cohort, the non-
542 transmitted PGSs could not be assessed, as no parental genotypes were available in the clinical
543 cohort. Additional validation in cohorts with complete omics data that applied the same
544 metabolomics and DNA methylation platforms are of large interest, but currently do not exist. Our
545 design is optimal in nearly all other aspects. The twin and clinical cohorts collected all data, including
546 biomarkers, at the same time, using the same arrays and platforms and with similar protocols, thus
547 decreasing unwanted sources of heterogeneity that are often present in sequential multi-omics
548 designs (Wörheide et al. 2021).

549 By adding a fourth broad exposome block, capturing known risk factors for childhood
550 aggression, such as neighborhood variables (Miller and Tolan 2019), to the multi-omics model, the
551 influence of environmental influences can be explored. Inclusion of other omics layers, such as the
552 transcriptome, proteome, or a microbiome, may give other insights into the biological mechanisms
553 of complex traits like childhood aggression. It should be noted that with the current method,
554 inclusion of additional omics layers will cause an increase in the computational burden. Therefore,
555 reduction of the computational burden might need to be considered in future studies. Parameter
556 reduction may be achieved by including a smaller number of CpGs, by for example including only the
557 top 1% or 5% most variable probes, or by selecting CpGs based on their known association with
558 aggressive behavior, such as the top CpGs from the EWAS meta-analysis for aggressive behavior (van
559 Dongen et al. 2021). While these suggestions, as well as increasing the sample size of training
560 datasets, will benefit future multi-omics investigations into childhood aggression, one of the first
561 steps in aggression research also concerns increasing the sample size of future genome-wide
562 association studies. To understand the underlying biological mechanisms of complex traits we
563 should not put all our efforts into a single approach. By exploring and optimizing various
564 methodologies and merging information from these different approaches we stand the best chance
565 of identifying robust mechanisms.

566 Conclusions

567 Our work entails one of the first applications of multi-omics approaches to childhood
568 psychopathology. The approach we used was developed for dichotomous traits and classification
569 purposes but also gives insight into the how different omics blocks associate with each other.
570 Classification was poor, whereas the multi-omics associations confirm well known associations
571 between childhood aggression and known risk factors as well as provide novel insight into the
572 correlational structure among omics variables from different omics blocks.

Declarations

Acknowledgments

The Netherlands Twin Register (NTR) warmly thanks all twin families for their participation. LUMC-Curium thanks all patients and their parents for participating, and clinicians for their support. NTR and LUMC-Curium are grateful to all researchers involved in the data collection for the ACTION Biomarker Study.

Funding

The current work is supported by the "Aggression in Children: Unraveling gene-environment interplay to inform Treatment and InterventiON strategies" project (ACTION) and the Consortium on Individual Development (CID). ACTION received funding from the European Union Seventh Framework Program (FP7/2007-2013) under grant agreement no 602768. CID is funded through the Gravitation Program of the Dutch Ministry of Education, Culture, and Science and the Netherlands Organization for Scientific Research (NWO grant number 024-001-003). The Netherlands Twin Register is supported by multiple grants from the Netherlands Organizations for Scientific Research (NWO) and Medical Research (ZonMW): Netherlands Twin Registry Repository (NWO 480-15-001/674); Why some children thrive (OCW Gravity program, NWO-024.001.003); Genetic influences on stability and change in psychopathology from childhood to young adulthood (ZonMw 912-10-020); Twin family database for behavior genomics studies (NWO 480-04-004); Twin research focusing on behavior (NWO 400-05-717); Longitudinal data collection from teachers of Dutch twins and their siblings (NWO 481-08-011), Twin-family-study of individual differences in school achievement (NWO-FES, 056-32-010), Genotype/phenotype database for behavior genetic and genetic epidemiological studies (ZonMw Middelgroot 911-09-032); the Biobank-based integrative omics study (BIOS) funded by BBMRI-NL (NWO projects 184.021.007 and 184.033.111); the European Science Council (ERC) Genetics of Mental Illness (ERC Advanced, 230374, PI Boomsma); Developmental trajectories of psychopathology (NIMH 1RC2 MH089995); the Avera Institute for Human Genetics, Sioux Falls, USA; the Royal Netherlands

598 Academy of Science Professor Award (PAH/6635) to D.I.B.. A research visit to the University of
599 Toulouse (France) by F.A.H. was supported by the Faculty of Behavioural and Movement Sciences
600 (FGB) Talent Fund (2019). J.D. is supported by the NWO-funded X-omics project (184.034.019). M.B. is
601 supported by an ERC consolidator grant (WELL-BEING 771057, PI Bartels).

602 Conflict of Interest

603 C.K. was employed by Good Biomarker Sciences (Leiden, the Netherlands). E.A.E. was employed
604 by the Avera Institute for Human Genetics (Sioux Falls, SD, United States). The remaining authors
605 declare that the research was conducted in the absence of any commercial or financial relationships
606 that could be construed as a potential conflict of interest.

607 Ethical approval

608 The study was conducted according to the guidelines of the Declaration of Helsinki, and approved
609 by the Central Ethics Committee on Research Involving Human Subjects of the VU University Medical
610 Center, Amsterdam (NTR 25th of May 2007 and ACTION 2013/41 and 2014.252), an Institutional
611 Review Board certified by the U.S. Office of Human Research Protections (IRB number IRB00002991
612 under Federal-wide Assurance- FWA00017598; IRB/institute codes), and the Medical Ethical
613 Committee of Leiden University Medical Center (B17.031, B17.032 and B17.040).

614 Informed consent

615 Parents provided written informed consent for their children and twin parents provided written
616 informed consent for their own participation.

617 Consent for publication

618 Not applicable.

619 Availability of data

620 The standardized protocol for large scale collection of urine and buccal-cell samples in the home
621 situation as developed for the ACTION Biomarker Study in children available at http://www.action-

622 euproject.eu/content/data-protocols. The data of the Netherlands Twin Register (NTR) ACTION
623 Biomarker Study may be accessed, upon approval of the data access committee, through the NTR
624 (https://tweelingenregister.vu.nl/information_for_researchers/working-with-ntr-data).

625 Code availability

626 Analysis code is available upon request from the corresponding author.

627 Authors' contributions

628 Conceptualization, F.A.H., J.D., M.B., S.D. and D.I.B.; methodology, S.D.; formal analysis, F.A.H.,
629 R.P. and J.J.H.; investigation, F.A.H., P.J.R., O.F.C., R.R.J.M.V., M.B. and D.I.B.; resources, A.C.H., C.K.,
630 E.A.E. and T.H.; data curation, C.E.M.B.; writing—original draft preparation, F.A.H., J.D., R.P., M.B.
631 and D.I.B..; writing—review and editing, all authors; visualization, F.A.H.; supervision, J.D., M.B. and
632 D.I.B.; funding acquisition, A.C.H., C.K., O.F.C., V.F., T.H., R.R.J.M.V., M.B. and D.I.B. All authors have
633 read and agreed to the published version of the manuscript.

634 ## References

635 Achenbach TM, Ivanova MY, Rescorla LA (2017) Empirically based assessment and taxonomy of
636 psychopathology for ages 1½–90+ years: Developmental, multi-informant, and multicultural
637 findings. Compr Psychiatry 79:4–18. https://doi.org/10.1016/j.comppsych.2017.03.006

638 Anderson CA, Bushman BJ (2002) Human Aggression. Annu Rev Psychol 53:27–51.
639 https://doi.org/10.1146/annurev.psych.53.100901.135231

640 Barban N, Jansen R, de Vlaming R, et al (2016) Genome-wide analysis identifies 12 loci influencing
641 human reproductive behavior. Nat Genet 48:1462–1472. https://doi.org/10.1038/ng.3698

642 Bartels M, Hendriks A, Mauri M, et al (2018) Childhood aggression and the co-occurrence of
643 behavioural and emotional problems: results across ages 3–16 years from multiple raters in six
644 cohorts in the EU-ACTION project. Eur Child Adolesc Psychiatry 27:1105–1121.
645 https://doi.org/10.1007/s00787-018-1169-1

646 Baselmans BML, Jansen R, Ip HF, et al (2019) Multivariate genome-wide analyses of the well-being
647 spectrum. Nat Genet 51:445–451. https://doi.org/10.1038/s41588-018-0320-8

648 Bates TC, Maher BS, Medland SE, et al (2018) The Nature of Nurture: Using a Virtual-Parent Design
649 to Test Parenting Effects on Children's Educational Attainment in Genotyped Families. Twin Res
650 Hum Genet 21:73–83. https://doi.org/10.1017/thg.2018.11

651 Beck JJ, Hottenga J-J, Mbarek H, et al (2019) Genetic Similarity Assessment of Twin-Family
652 Populations by Custom-Designed Genotyping Array. Twin Res Hum Genet 22:210–219.
653 https://doi.org/10.1017/thg.2019.41

654 Bellair PE, McNulty TL, Piquero AR (2019) Persistent material hardship and childhood physical
655 aggression. Aggress Violent Behav 49:101309. https://doi.org/10.1016/j.avb.2019.07.004

656 Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful

657 approach to multiple testing. J. R. Stat. Soc. B 57:289–300

658 Benyamin B, Pourcain B, Davis OS, et al (2014) Childhood intelligence is heritable, highly polygenic
659 and associated with FNBP1L. Mol Psychiatry 19:253–258.
660 https://doi.org/10.1038/mp.2012.184

661 Bhak Y, Jeong H, Cho YS, et al (2019) Depression and suicide risk prediction models using blood-
662 derived multi-omics data. Transl Psychiatry 9:262. https://doi.org/10.1038/s41398-019-0595-2

663 Boomsma DI (2015) Aggression in children: unravelling the interplay of genes and environment
664 through (epi) genetics and metabolomics. J Pediatr Neonatal Individ Med 4:e040251.
665 https://doi.org/10.7363/040251

666 Branje S, Geeraerts S, de Zeeuw EL, et al (2020) Intergenerational transmission: Theoretical and
667 methodological issues and an introduction to four Dutch cohorts. Dev Cogn Neurosci
668 45:100835. https://doi.org/10.1016/j.dcn.2020.100835

669 Braun PR, Tanaka-Sahker M, Chan AC, et al (2019) Genome-wide DNA methylation investigation of
670 glucocorticoid exposure within buccal samples. Psychiatry Clin Neurosci 73:323–330.
671 https://doi.org/10.1111/pcn.12835

672 Cecil CAM, Walton E, Jaffee SR, et al (2018a) Neonatal DNA methylation and early-onset conduct
673 problems: A genome-wide, prospective study. Dev Psychopathol 30:383–397.
674 https://doi.org/10.1017/S095457941700092X

675 Cecil CAM, Walton E, Pingault JB, et al (2018b) DRD4 methylation as a potential biomarker for
676 physical aggression: An epigenome-wide, cross-tissue investigation. Am J Med Genet Part B
677 Neuropsychiatr Genet 177:746–764. https://doi.org/10.1002/ajmg.b.32689

678 Chen X, Xu J, Tang J, et al (2020) Dysregulation of amino acids and lipids metabolism in schizophrenia
679 with violence. BMC Psychiatry 20:1–11. https://doi.org/10.1186/s12888-020-02499-y

680 Dawiskiba T, Deja S, Mulak A, et al (2014) Serum and urine metabolomic fngerprinting in diagnostics
681 of inflammatory bowel diseases. World J Gastroenterol 20:163–174.
682 https://doi.org/10.3748/wjg.v20.i1.163

683 Dean KR, Hammamieh R, Mellon SH, et al (2019) Multi-omic biomarker identification and validation
684 for diagnosing warzone-related post-traumatic stress disorder. Mol Psychiatry.
685 https://doi.org/10.1038/s41380-019-0496-z

686 Demontis D, Walters RK, Martin J, et al (2019) Discovery of the first genome-wide significant risk loci
687 for attention deficit/hyperactivity disorder. Nat Genet 51:63–75.
688 https://doi.org/10.1038/s41588-018-0269-7

689 Duruflé H, Selmani M, Ranocha P, et al (2020) A powerful framework for an integrative study with
690 heterogeneous omics data: from univariate statistics to multi-block analysis. Brief Bioinform
691 00:1–13. https://doi.org/10.1093/bib/bbaa166

692 Ehli EA, Abdellaoui A, Fedko IO, et al (2017) A method to customize population-specific arrays for
693 genome-wide association testing. Eur J Hum Genet 25:267–270.
694 https://doi.org/10.1038/ejhg.2016.152

695 Fergusson DM, Horwood LJ, Ridder EM (2005) Show me the child at seven: The consequences of
696 conduct problems in childhood for psychosocial functioning in adulthood. J Child Psychol
697 Psychiatry Allied Discip 46:837–849. https://doi.org/10.1111/j.1469-7610.2004.00387.x

698 Galili T (2015) dendextend: An R package for visualizing, adjusting and comparing trees of
699 hierarchical clustering. Bioinformatics 31:3718–3720.
700 https://doi.org/10.1093/bioinformatics/btv428

701 Glad CAM, Andersson-Assarsson JC, Berglund P, et al (2017) Reduced DNA methylation and
702 psychopathology following endogenous hypercortisolism - A genome-wide study. Sci Rep 7:1–
703 11. https://doi.org/10.1038/srep44445

González I, Cao K-AL, Davis MJ, Déjean S (2012) Visualising associations between paired 'omics' data sets. BioData Min 5:19. https://doi.org/10.1186/1756-0381-5-19

Grove J, Ripke S, Als TD, et al (2019) Identification of common genetic risk variants for autism spectrum disorder. Nat Genet 51:431–444. https://doi.org/10.1038/s41588-019-0344-8

Guillemin C, Provençal N, Suderman M, et al (2014) DNA Methylation Signature of Childhood Chronic Physical Aggression in T Cells of Both Men and Women. PLoS One 9:e86822. https://doi.org/10.1371/journal.pone.0086822

Gulsun M, Oznur T, Aydemir E, et al (2016) Possible relationship between amino acids, aggression and psychopathy. Int J Psychiatry Clin Pract 1501:1–10. https://doi.org/10.3109/13651501.2016.1144771

Hagenbeek FA, Kluft C, Hankemeier T, et al (2016) Discovery of biochemical biomarkers for aggression: A role for metabolomics in psychiatry. Am J Med Genet Part B Neuropsychiatr Genet 171:719–732. https://doi.org/10.1002/ajmg.b.32435

Hagenbeek FA, Roetman PJ, Pool R, et al (2020) Urinary Amine and Organic Acid Metabolites Evaluated as Markers for Childhood Aggression: The ACTION Biomarker Study. Front Psychiatry 11:. https://doi.org/10.3389/fpsyt.2020.00165

Hendriks AM, Finkenauer C, Nivard MG, et al (2020) Comparing the genetic architecture of childhood behavioral problems across socioeconomic strata in the Netherlands and the United Kingdom. Eur Child Adolesc Psychiatry 29:353–362. https://doi.org/10.1007/s00787-019-01357-x

Hompes T, Izzi B, Gellens E, et al (2013) Investigating the influence of maternal cortisol and emotional state during pregnancy on the DNA methylation status of the glucocorticoid receptor gene (NR3C1) promoter region in cord blood. J Psychiatr Res 47:880–891. https://doi.org/10.1016/j.jpsychires.2013.03.009

Ip HF, van der Laan CM, Krapohl EML, et al (2021) Genetic association study of childhood aggression

728 across raters, instruments, and age. Transl Psychiatry 11:413. https://doi.org/10.1038/s41398-
729 021-01480-x

730 Jakovljevic M, Jakovljevic I (2019) A Transdisciplinary Integrative Approach for Precision Psychiatry.
731 In: Kim Y-K (ed) Frontiers in Psychiatry. Springer Singapore, Singapore, pp 399–428

732 Jansen PR, Watanabe K, Stringer S, et al (2019) Genome-wide analysis of insomnia in 1,331,010
733 individuals identifies new risk loci and functional pathways. Nat Genet 51:394–403.
734 https://doi.org/10.1038/s41588-018-0333-3

735 Jiang S, Postovit L, Cattaneo A, et al (2019) Epigenetic Modifications in Stress Response Genes
736 Associated With Childhood Trauma. Front Psychiatry 10:1–19.
737 https://doi.org/10.3389/fpsyt.2019.00808

738 Kassing F, Godwin J, Lochman JE, Coie JD (2019) Using Early Childhood Behavior Problems to Predict
739 Adult Convictions. J Abnorm Child Psychol 47:765–778. https://doi.org/10.1007/s10802-018-
740 0478-7

741 Kong A, Thorleifsson G, Frigge ML, et al (2018) The nature of nurture: Effects of parental genotypes.
742 Science (80-) 359:424–428. https://doi.org/10.1126/science.aan6877

743 Lê Cao KA, Boitard S, Besse P (2011) Sparse PLS discriminant analysis: Biologically relevant feature
744 selection and graphical displays for multiclass problems. BMC Bioinformatics 12:.
745 https://doi.org/10.1186/1471-2105-12-253

746 Lee JJ, Wedow R, Okbay A, et al (2018) Gene discovery and polygenic prediction from a genome-
747 wide association study of educational attainment in 1.1 million individuals. Nat Genet 50:1112–
748 1121. https://doi.org/10.1038/s41588-018-0147-3

749 Ligthart L, van Beijsterveldt CEM, Kevenaar ST, et al (2019) The Netherlands Twin Register:
750 Longitudinal Research Based on Twin and Twin-Family Designs. Twin Res Hum Genet 22:623–
751 636. https://doi.org/10.1017/thg.2019.93

752 Lima AR, Pinto J, Amaro F, et al (2021) Advances and Perspectives in Prostate Cancer Biomarker
753 Discovery in the Last 5 Years through Tissue and Urine Metabolomics. Metabolites 11:181.
754 https://doi.org/10.3390/metabo11030181

755 Liu J (2004) Malnutrition at Age 3 Years and Externalizing Behavior Problems at Ages 8, 11, and 17
756 Years. Am J Psychiatry 161:2005–2013. https://doi.org/10.1176/appi.ajp.161.11.2005

757 Liu M, Jiang Y, Wedow R, et al (2019) Association studies of up to 1.2 million individuals yield new
758 insights into the genetic etiology of tobacco and alcohol use. Nat. Genet. 51:237–244

759 Martínez-Reyes I, Chandel NS (2020) Mitochondrial TCA cycle metabolites control physiology and
760 disease. Nat Commun 11:1–11. https://doi.org/10.1038/s41467-019-13668-3

761 Masud H, Ahmad MS, Cho KW, Fakhr Z (2019) Parenting Styles and Aggression Among Young
762 Adolescents: A Systematic Review of Literature. Community Ment Health J 55:1015–1030.
763 https://doi.org/10.1007/s10597-019-00400-0

764 Meng Y, Groth SW (2018) Fathers Count: The Impact of Paternal Risk Factors on Birth Outcomes.
765 Matern Child Health J 22:401–408. https://doi.org/10.1007/s10995-017-2407-8

766 Miller GM, Tolan PH (2019) The influence of parenting practices and neighborhood characteristics on
767 the development of childhood aggression. J Community Psychol 47:135–146.
768 https://doi.org/10.1002/jcop.22105

769 Mitjans M, Seidel J, Begemann M, et al (2019) Violent aggression predicted by multiple pre-adult
770 environmental hits. Mol Psychiatry 24:1549–1564. https://doi.org/10.1038/s41380-018-0043-3

771 Moran S, Arribas C, Esteller M (2016) Validation of a DNA methylation microarray for 850,000 CpG
772 sites of the human genome enriched in enhancer sequences. Epigenomics 8:389–399.
773 https://doi.org/10.2217/epi.15.114

774 Mulligan CJ, D'Errico NC, Stees J, Hughes DA (2012) Methylation changes at NR3C1 in newborns

775 associate with maternal prenatal stress exposure and newborn birth weight. Epigenetics
776 7:853–857. https://doi.org/10.4161/epi.21180

777 Odintsova V V., Roetman PJ, Ip HF, et al (2019) Genomics of human aggression. Psychiatr Genet
778 29:170–190. https://doi.org/10.1097/YPG.0000000000000239

779 Ooi M, Nishiumi S, Yoshie T, et al (2011) GC/MS-based profiling of amino acids and TCA cycle-related
780 molecules in ulcerative colitis. Inflamm Res 60:831–840. https://doi.org/10.1007/s00011-011-
781 0340-7

782 Pinu FR, Beale DJ, Paten AM, et al (2019) Systems biology and multi-omics integration: Viewpoints
783 from the metabolomics research community. Metabolites 9:1–31.
784 https://doi.org/10.3390/metabo9040076

785 R Core Team R: A language and environment for statistical computing. https://www.r-project.org/

786 Radwan K, Coccaro EF (2020) Comorbidity of disruptive behavior disorders and intermittent
787 explosive disorder. Child Adolesc Psychiatry Ment Health 14:24.
788 https://doi.org/10.1186/s13034-020-00330-w

789 Rist MJ, Roth A, Frommherz L, et al (2017) Metabolite patterns predicting sex and age in participants
790 of the Karlsruhe Metabolomics and Nutrition (KarMeN) study. PLoS One 12:1–21.
791 https://doi.org/10.1371/journal.pone.0183228

792 Rogers P, Stoner J (2016) Modification of the Sandwich Estimator in Generalized Estimating
793 Equations with Correlated Binary Outcomes in Rare Event and Small Sample Settings. Am J Appl
794 Math Stat 3:243–251. https://doi.org/110.1016/j.bbi.2017.04.008

795 Rohart F, Gautier B, Singh A, Lê Cao KA (2017) mixOmics: An R package for 'omics feature selection
796 and multiple data integration. PLoS Comput Biol 13:1–20.
797 https://doi.org/10.1371/journal.pcbi.1005752

798 Savage JE, Jansen PR, Stringer S, et al (2018) Genome-wide association meta-analysis in 269,867
799 individuals identifies new genetic and functional links to intelligence. Nat Genet 50:912–919.
800 https://doi.org/10.1038/s41588-018-0152-6

801 Schicho R, Shaykhutdinov R, Ngo J, et al (2012) Quantitative metabolomic profiling of serum, plasma,
802 and urine by 1H NMR spectroscopy discriminates between patients with inflammatory bowel
803 disease and healthy individuals. J Proteome Res 11:3344–3357.
804 https://doi.org/10.1021/pr300139q

805 Selhub J (1999) Homocystein Metabolism. Annu Rev Nutr 19:217–246.
806 https://doi.org/10.1146/annurev.nutr.19.1.217

807 Siever LJ (2008) Neurobiology of Aggression and Violence. Am J Psychiatry 165:429–442

808 Singh A, Shannon CP, Gautier B, et al (2019) DIABLO: An integrative approach for identifying key
809 molecular drivers from multi-omics assays. Bioinformatics 35:3055–3062.
810 https://doi.org/10.1093/bioinformatics/bty1054

811 Sinke L, van Iterson M, Cats D, et al DNAmArray: Streamlined workflow for the quality control,
812 normalization, and analysis of Illumina methylation array data.
813 https://molepi.github.io/DNAmArray_workflow/

814 Subramanian I, Verma S, Kumar S, et al (2020) Multi-omics Data Integration, Interpretation, and Its
815 Application. Bioinform Biol Insights 14:117793221989905.
816 https://doi.org/10.1177/1177932219899051

817 Tibshirani R (1996) Regression Shrinkage and Selection via the Lasso. J R Stat Soc Ser B (Methodol)
818 58:267–288

819 Van Adrichem DS, Huijbregts SCJ, Van Der Heijden KB, et al (2020) Aggressive behavior during
820 toddlerhood: Interrelated effects of prenatal risk factors, negative affect, and cognition. Child
821 Neuropsychol 26:982–1004. https://doi.org/10.1080/09297049.2020.1769582

822 van der Laan CM, Morosoli-García JJ, van de Weijer SGA, et al (2021) Continuity of Genetic Risk for
823 Aggressive Behavior Across the Life-Course. Behav Genet. https://doi.org/10.1007/s10519-021-
824 10076-6

825 van Dongen J, Hagenbeek FA, Suderman M, et al (2021) DNA methylation signatures of aggression
826 and closely related constructs: A meta-analysis of epigenome-wide studies across the lifespan.
827 Mol Psychiatry 2020.07.22.215939. https://doi.org/10.1038/s41380-020-00987-x

828 Vaughn M, Salas-Wright C, Naeger S, et al (2016) Childhood Reports of Food Neglect and Impulse
829 Control Problems and Violence in Adulthood. Int J Environ Res Public Health 13:389.
830 https://doi.org/10.3390/ijerph13040389

831 Vuoksimaa E, Rose RJ, Pulkkinen L, et al (2020) Higher aggression is related to poorer academic
832 performance in compulsory education. J Child Psychol Psychiatry.
833 https://doi.org/10.1111/jcpp.13273

834 Watanabe K, Stringer S, Frei O, et al (2019) A global overview of pleiotropy and genetic architecture
835 in complex traits. Nat Genet 51:1339–1348. https://doi.org/10.1038/s41588-019-0481-0

836 Weaver ICG, Cervoni N, Champagne FA, et al (2004) Epigenetic programming by maternal behavior.
837 Nat Neurosci 7:847–854. https://doi.org/10.1038/nn1276

838 Whipp AM, Korhonen T, Raevuori A, et al (2019) Early adolescent aggression predicts antisocial
839 personality disorder in young adults: a population-based study. Eur Child Adolesc Psychiatry
840 28:341–350. https://doi.org/10.1007/s00787-018-1198-9

841 Whipp AM, Vuoksimaa E, Bolhuis K, et al (2021a) Teacher-rated aggression and co-occurring
842 behaviors and emotional problems among schoolchildren in four population-based European
843 cohorts. PLoS One 16:e0238667. https://doi.org/10.1371/journal.pone.0238667

844 Whipp AM, Vuoksimaa E, Korhonen T, et al (2021b) Ketone body 3-hydroxybutyrate as a biomarker
845 of aggression. Sci Rep 11:5813. https://doi.org/10.1038/s41598-021-84635-6

846 Wörheide MA, Krumsiek J, Kastenmüller G, Arnold M (2021) Multi-omics integration in biomedical
847 research – A metabolomics-centric review. Anal Chim Acta 1141:144–162.
848 https://doi.org/10.1016/j.aca.2020.10.038

849 Wray NR, Ripke S, Mattheisen M, et al (2018) Genome-wide association analyses identify 44 risk
850 variants and refine the genetic architecture of major depression. Nat Genet 50:668–681.
851 https://doi.org/10.1038/s41588-018-0090-3

852 Wrigglesworth J, Ancelin ML, Ritchie K, Ryan J (2019) Association between DNA methylation of the
853 KITLG gene and cortisol levels under stress: a replication study. Stress 22:162–168.
854 https://doi.org/10.1080/10253890.2018.1519019

855 Wu Y, Zhong X, Lin Y, et al (2021) Estimating genetic nurture with summary statistics of
856 multigenerational genome-wide association studies. Proc Natl Acad Sci 118:e2023184118.
857 https://doi.org/10.1073/pnas.2023184118

858 Xiong Z, Yang F, Li M, et al (2022) EWAS Open Platform: integrated data, knowledge and toolkit for
859 epigenome-wide association study. Nucleic Acids Res 50:D1004–D1009.
860 https://doi.org/10.1093/nar/gkab972

861 Yuan Y, Zhao J, Li T, et al (2021) Integrative metabolic profile of myelodysplastic syndrome based on
862 UHPLC–MS. Biomed Chromatogr 1–12. https://doi.org/10.1002/bmc.5136

863

Figures

Fig. 1. Overview of the biomarker identification approach for childhood aggression—details of statistical analyses and data included in each analysis.

We employed an analytical design comprising three phases: 1) single-omics analyses; 2) pairwise cross-omics analyses; and 3) multi-omics analyses. First, we performed univariate polygenic score (PGS) analysis in 70% of the twin data and built multivariate single-omics biomarker panels in the twin cohort, with 70% of the twin data for model training (training data), 30% of the twin data and the clinical cohort for model testing (test data). Second, we examined the overall pairwise cross-omics correlations and the pairwise correlations of those omics variables selected by the single-omics models in the training data. Third, using the same data, we compared three multi-omics models, with different assumptions of the correlations among the omics blocks, and describe the multi-omics relationships of the selected omics variables. We offer the analytical details in the **Methods** section.

Study Phase	Analyses	Design
1. Single-omics	Univariate: • Polygenic score analysis	Training data: • 70% twin cohort (*N* = 645) • *Cases*: 271 (42.0%) • *Controls*: 374 (58.0%)
	Multivariate – model training: • Principal Component Analysis (PCA) • Partial Least Squares Discriminant Analysis (PLS-DA) • Sparse Partial Least Squares Discriminant Analysis (sPLS-DA)	
	Multivariate – model testing: • Predict out-of-sample case-control status • Receiver Operating Characteristic (ROC) analysis	Test data: • 30% twin cohort (*N* = 277) • *Cases:* 117 (42.2%) • *Controls:* 160 (57.8%) Clinical data: • Clinical cohort (*N* = 142) • *Cases:* 64 (45.1%) • *Controls:* 78 (54.9%)
2. Pairwise cross-omics	All omics variables: • Partial Least Squares (PLS) analysis Selected omics variables: • Partial Least Squares (PLS) analysis • Hierarchical clustering	Training data: • 70% twin cohort (*N* = 645)
3. Multi-omics	Model training: • Multi-block Partial Least Squares Discriminant Analysis (MB-PLS-DA) • Multi-block sparse Partial Least Squares Discriminant Analysis (MB-sPLS-DA)	Training data: • 70% twin cohort (*N* = 645) • *Cases*: 271 (42.0%) • *Controls*: 374 (58.0%)
	Model testing: • Predict out-of-sample case-control status • Receiver Operating Characteristic (ROC) analysis	Test data: • 30% twin cohort (*N* = 277) • *Cases:* 117 (42.2%) • *Controls:* 160 (57.8%) Clinical data: • Clinical cohort (*N* = 142) • *Cases:* 64 (45.1%) • *Controls:* 78 (54.9%)

877

878 **Fig. 2.** Clustered heatmaps of the relationships obtained by the pairwise cross-omics Partial Least
879 Squares (PLS) regression models including only the omics variables as selected by the single-omics
880 sparse Partial Least Squares Discriminant Analyses (sPLS-DA).

881 We generated the hierarchical clustering using the Ward linkage algorithm on Euclidean distances of
882 the PLS variates. For each dendrogram we identified two clusters (cluster 1 = pink, cluster 2 = blue).
883 We have depicted positive relationships among the omics variables in red and negative relationships
884 in blue. For the polygenic scores (PGSs), the '_NTm' suffix denotes non-transmitted maternal PGSs,
885 the '_NTf' suffix denotes the non-transmitted paternal PGSs, and childhood aggression is
886 abbreviated as "aggression", Attention-Deficit Hyperactivity Disorder as "ADHD", Major Depressive
887 Disorder as "MDD", Autism Spectrum Disorder as "Autism", Educational Attainment as "EA", and
888 wellbeing spectrum as "wellbeing". For the metabolites, the 'amines.' prefix shows we measured
889 these metabolites on the Liquid Chromatography Mass Spectrometry (LC-MS) amines platform, the
890 'steroids.' prefix shows we measured these metabolites on the LC-MS steroids platform, and the
891 'OA.' prefix shows we measured these metabolites on the Gas Chromatography (GC-) MS organic
892 acids platform. **(a)** Relationships of the 1,614 CpGs and 90 metabolites included in the 3-component
893 DNA methylation-metabolomics PLS model, where the selected CpGs are represented in the
894 columns and the metabolomics traits in the rows. We included the cluster assignments and the full
895 data matrix in **Data S4-S5**. **(b)** Relationships of the 36 PGSs and 90 metabolites included in the 2-
896 component PGSs-metabolomics PLS model, where the selected PGSs are represented in the columns
897 and the metabolomics traits in the rows. We included the cluster assignments and the full data
898 matrix in **Table S9** and **Data S6**, respectively. **(c)** Relationships of the 36 PGSs and 1,614 CpGs
899 included in the 2-component PGSs-DNA methylation PLS model, where the selected PGSs are
900 represented in the columns and the CpGs in the rows. We included the cluster assignments and the
901 full data matrix in **Data S7-S8**.

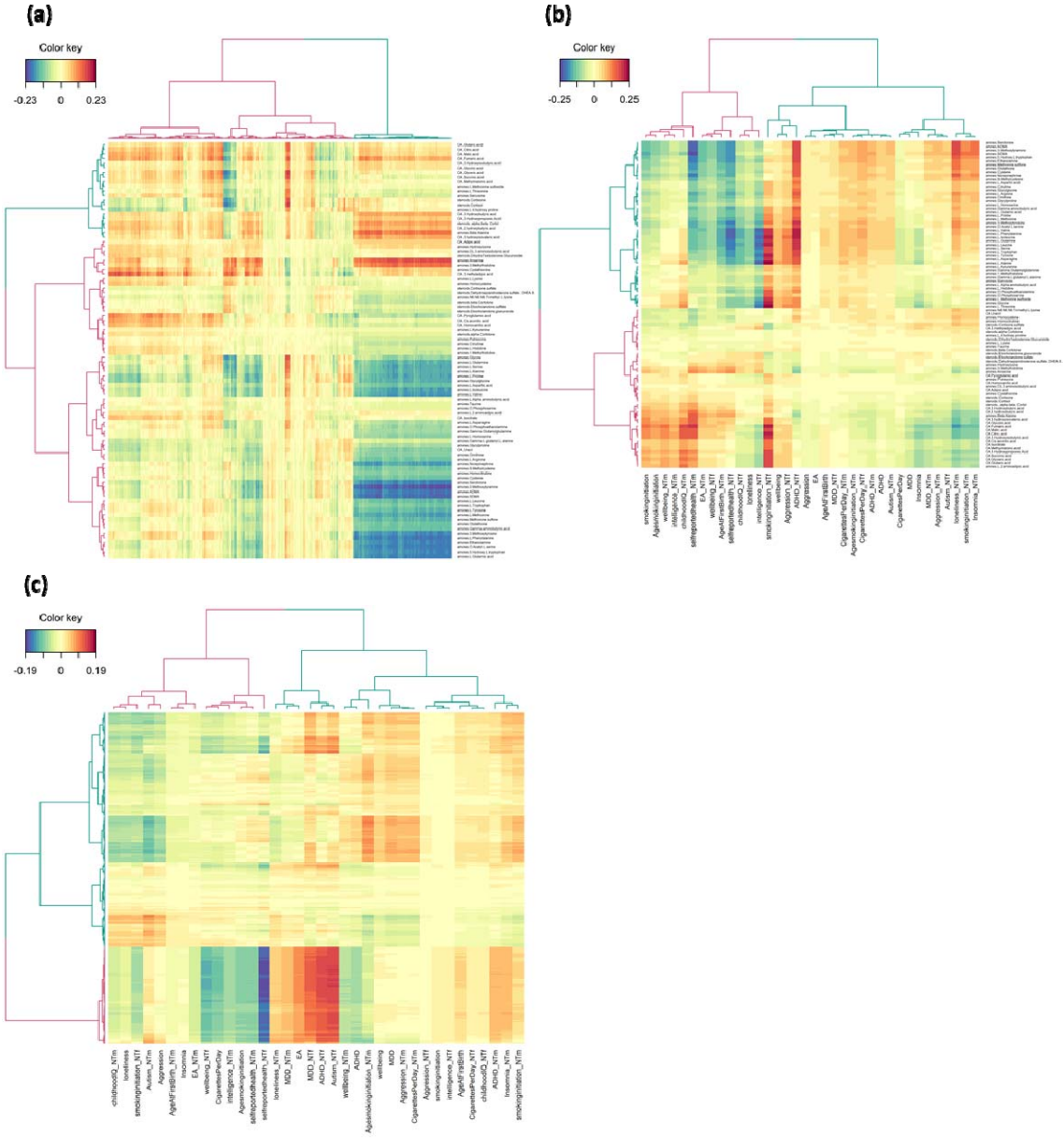
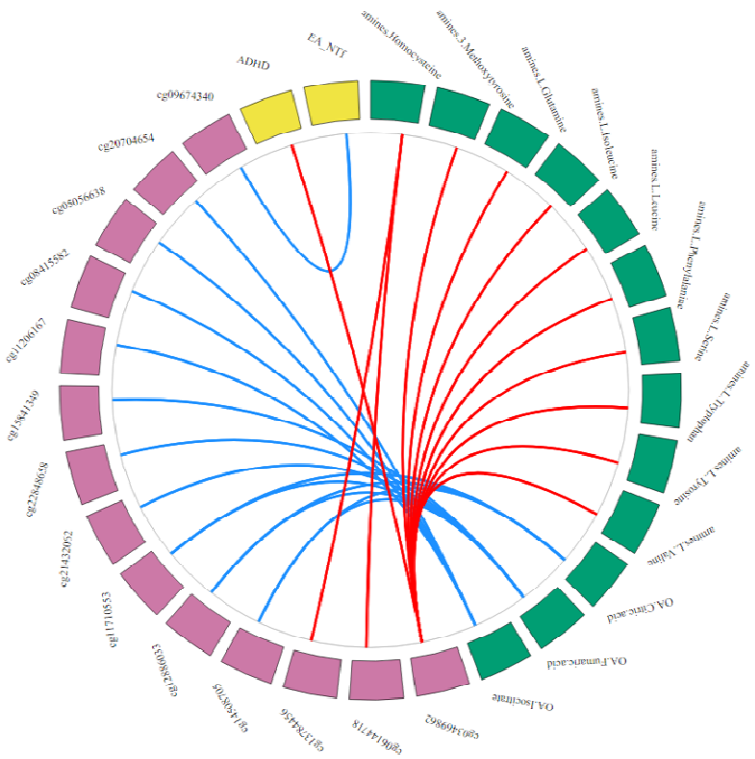

(a)
Color key
-0.23
0
0.23
(b)
Color key
-0.25
0
0.25
(c)
Color key
-0.19
0
0.19

Fig. 3. Strong cross-omics connections of the multi-omics traits identified in the 5-component multi-block sparse Partial Least Squares Discriminant Analysis (MB-sPLS-DA) model including the empirical design matrix

The outer ring depicts the PGSs, CpGs, and metabolites in yellow, pink, and green, respectively. For the polygenic scores (PGSs), Attention-Deficit Hyperactivity Disorder is abbreviated as "ADHD", and Educational Attainment as "EA", and the '_NTf' suffix denotes the non-transmitted paternal PGSs. For the metabolites, the 'amines.' prefix shows we measured these metabolites on the Liquid Chromatography Mass Spectrometry (LC-MS) amines platform, and the 'OA.' prefix shows we measured these metabolites on the Gas Chromatography (GC-) MS organic acids platform. The inner plot depicts the connetions among the omics variables. Here, we depict only high absolute correlations of the PLS variates ($|r| \geq 0.60$) between variables of at least two omics blocks, with blue lines reflecting negative correlations and red lines positive correlations. We averaged correlations of the PLS variates across all components in the MB-sPLS-DA model. We included the full data matrix in **Data S10** and the patterns in **Table S15**.



Tables

Table 1. Overview of the discovery genome-wide association studies to calculate polygenetic scores.

Trait [1]	N discovery GWA	Reference	Source original summary statistics
Childhood aggression	151,741	(Ip et al. 2021)	Paper accepted, summary statistics obtained from authors
Attention-Deficit Hyperactivity Disorder	20,183 cases & 35,191 controls	(Demontis et al. 2019)	http://ipsych.au.dk/downloads/data-download-agreement-adhd-european-ancestry-gwas-june-2017
Major Depressive Disorder	135,458 cases & 344,901 controls	(Wray et al. 2018)	https://www.med.unc.edu/pgc/results-and-downloads/
Autism Spectrum Disorder	18,381 cases & 27,969 controls	(Grove et al. 2019)	https://www.med.unc.edu/pgc/results-and-downloads/
Loneliness	355,583	http://www.nealelab.is/uk-biobank/	https://www.dropbox.com/s/nf4jl3mdppu1ng8/2020.gwas.imputed_v3.both_sexes.tsv.bgz?dl=0
Insomnia	1,331,010	(Jansen et al. 2019)	https://ctg.cncr.nl/documents/p1651/Insomnia_sumstats_Jansenetal.txt.gz
Self-reported Health	359,681	http://www.nealelab.is/uk-biobank/	https://www.dropbox.com/s/aawh07hlhldbckc/2178.gwas.imputed_v3.both_sexes.tsv.bgz?dl=0
Smoking initiation (Ever/never smoked)	1,232,091	(Liu et al. 2019)	https://conservancy.umn.edu/bitstream/handle/11299/201564/SmokingInitiation.txt.gz?sequence=27&isAllowed=y
Age of smoking initiation	94,891	(Watanabe et al. 2019)	https://atlas.ctglab.nl/ukb2_sumstats/f.2867.0.0_res.EUR.sumstats.MACfilt.txt.gz
Cigarettes per day	37,334	(Liu et al. 2019)	https://conservancy.umn.edu/bitstream/handle/11299/201564/CigarettesPerDay.txt.gz?sequence=17&isAllowed=y
Childhood IQ	17,989	(Benyamin et al. 2014)	http://ssgac.org/documents/CHIC_Summary_Benyamin2014.txt.gz
Educational Attainment[5]	1,131,881	(Lee et al. 2018)	https://www.dropbox.com/s/ho58e9jmytmpaf8/GWAS_EA_excl23andMe.txt?dl=0
Age at first birth	251,151	(Barban et al. 2016)	http://sociogenome.com/material/GWASresults/AgeFirstBirth_Pooled.txt.gz
Wellbeing spectrum	2,370,390	(Baselmans et al. 2019)	https://surfdrive.surf.nl/files/index.php/s/Ow1qCDpFT421ZOO/download?path=%2FMultivariate_GWAMA_sumstats%2FN_GWAMA&files=N_GWAMA_WBspectrum_no23andME.txt.gz
Intelligence	269,867	(Savage et al. 2018)	https://ctg.cncr.nl/documents/p1651/SavageJansen_IntMeta_sumstats.zip

[1] In the (supplementary) tables and figures childhood aggression is abbreviated as “aggression”, Attention-Deficit Hyperactivity Disorder as “ADHD”, Major Depressive Disorder as “MDD”, Autism Spectrum Disorder as “Autism”, Educational Attainment as “EA”, and wellbeing spectrum as “wellbeing”.

Table 2. Demographics of the training, test, and clinical data.

	Training data (70% twin cohort)			Test data (30% twin cohort)			Clinical data (Clinical cohort)		
	Controls	**Cases**	**Total**	**Controls**	**Cases**	**Total**	**Controls**	**Cases**	**Total**
N (%)	374 (58.0%)	271 (42.0%)	645 (100%)	160 (57.8%)	117 (42.2%)	277 (100%)	78 (54.9%)	64 (45.1%)	142 (100%)
N (%) complete twin pairs	128	80	293	56	35	128	-	-	-
Mean (*SD*) age	9.4 (1.9)	9.6 (1.8)	9.5 (1.9)	9.6 (1.9)	9.6 (1.9)	9.6 (1.9)	10.8 (1.7)	9.6 (1.7)	10.2 (1.8)
Range age	6.1 - 12.7	6.1 - 12.9	6.1 - 12.9	5.6 - 12.6	5.8 - 12.8	5.6 - 12.8	6.5 - 13.4	6.3 - 13.3	6.3 - 13.4
N (%) females	199 (53.2%)	116 (42.8%)	315 (48.8%)	81 (50.6%)	55 (47.0%)	136 (49.1%)	20 (25.6%)	19 (29.7%)	39 (27.5%)
N (%) MZ twins	299 (80.0%)	228 (84.1%)	527 (81.7%)	118 (73.7%)	108 (92.3%)	226 (81.6%)	-	-	-
Mean (*SD*) aggression score [1]	3.3 (4.1)	7.3 (5.8)	5.0 (5.3)	3.1 (4.0)	7.6 (6.6)	5.0 (5.7)	6.1 (3.4)	21.1 (4.8)	12.9 (8.5)

Notes: MZ, monozygotic.

[1] Measured with the mother-rated Aggressive Behavior syndrome scale of the Achenbach System of Empirically Based Assessment (ASEBA) Child Behavior Checklist (CBCL). The ASEBA CBCL Aggressive Behavior scores in the clinical cohort include 90% mother report and 10% father report.