Behavior Genetics (2023) 53:101–117
https://doi.org/10.1007/s10519-022-10126-7

ORIGINAL RESEARCH

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]

Received: 3 September 2021 / Accepted: 25 October 2022 / Published online: 7 November 2022

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, 78,772 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

Introduction

Omics studies can lead to an improved understanding of the biological mechanisms contributing to mental health and disorders (Jakovljevic and Jakovljevic 2019). Different omics technologies, e.g., genomics, epigenomics, transcriptomics, or metabolomics, assess different aspects of the development and progression of complex traits and disorders. The analysis of a single omics layer provides a unique, but likely incomplete, picture of the underlying biology, whereas studies combining multiple omics layers may lead to more comprehensive insights into human biology,

Edited by David Evans.

✉ Fiona A. Hagenbeek
f.a.hagenbeek@vu.nl

1 Department of Biological Psychology, Vrije Universiteit Amsterdam, Van der Boechorststraat 7-10, 1081 BT Amsterdam, The Netherlands

2 Amsterdam Public Health Research Institute, Amsterdam, The Netherlands

3 Amsterdam Reproduction & Development (AR&D) Research Institute, Amsterdam, The Netherlands

4 Department of Child and Adolescent Psychiatry, LUMC-Curium, 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

because the different omics layers are interrelated and interact (Wörheide et al. 2021). Multi-omics analyses can aid in biomarker discovery, diagnosis, patient classification or subtyping, evaluation of treatment response and uncovering novel insights into disease biology (Pinu et al. 2019; Subramanian et al. 2020).

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

Large-scale genome-wide association (GWA) studies on aggression have not yet identified significant single nucleotide polymorphisms (SNPs) (Odintsova et al. 2019; Ip et al. 2021). Gene-based analysis for childhood aggression yielded three significantly associated genes (*ST3GAL3*, *PCDH7*, and *IPO13*) (Ip et al. 2021). The polygenic score for childhood aggression does not only significantly explain childhood aggressive behavior at age 7 (Ip et al. 2021), but it also associated with aggression at age 12 to 41 in a Dutch sample and age 38 to 48 in an Australian sample (van der Laan et al. 2021). Small-scale epigenetic studies (sample size range: 41–260) showed that DNA methylation differences in various tissues are associated with aggression and related traits in children and adults (Guillemin et al. 2014; Cecil et al. 2018b, a; Mitjans et al. 2019). The first large-scale epigenome-wide association study (EWAS) meta-analysis across child and adult cohorts ($N = 15{,}324$) reported 13 significant sites in peripheral blood for broad aggression across the lifespan (van Dongen et al. 2021). Metabolomics studies (sample size range: 77–725) detected plasma and serum metabolites associated with aggression and related traits in adults (Gulsun et al. 2016; Chen et al. 2020; Whipp et al. 2021b). A study in 1347 twins and 183 clinical cases found significant associations of urinary metabolites with childhood aggression (Hagenbeek et al. 2020). These single omics approaches hint at potentially important biological pathways for human aggression, with most of these findings awaiting replication.

In this study, we aim to integrate multiple omics layers to construct a multi-omics biomarker panel for childhood aggressive behavior and explore the correlations among the omics traits included in this panel. We collected biological samples in a subproject of ACTION (Aggression in Children: Unraveling gene-environment interplay to inform Treatment and InterventiON strategies): the ACTION Biomarker Study (Boomsma 2015; Bartels et al. 2018; Hagenbeek et al. 2020). The ACTION Biomarker Study comprised a cohort of twins from the Netherlands Twin Register (NTR) (Ligthart et al. 2019) and a clinical cohort of children referred to a youth psychiatry clinic (LUMC-Curium, the Netherlands). The genome-wide SNP and DNA methylation (Illumina EPIC 850K array) data, and the urinary amines and organic acids as measured in these cohorts were previously included in a genome-wide genetic and epigenetic association study and a metabolomics study of (childhood) aggression (Hagenbeek et al. 2020; van Dongen et al. 2021; Ip et al. 2021). We expand on these single-omics studies by integrating these data with a third metabolomics dataset, focusing on urinary steroid hormones. We calculated transmitted and non-transmitted polygenic scores (PGSs) for childhood aggression and a series of genetically correlated traits, such as Attention-Deficit Hyperactivity Disorder (ADHD), smoking, and intelligence. The non-transmitted PGSs assess the indirect effects of genetic variants that were not transmitted from parents to offspring, i.e., genetic nurture, and capture the effects of the environment created by parents beyond the genetic intergeneration transmission (Bates et al. 2018; Kong et al. 2018). Thus, if alleles not transmitted from parents to offspring affect offspring outcomes, this indicates that the offspring's home-environment is influenced by parental genotypes, and the home-environment in turn affects offspring outcomes (Branje et al. 2020).

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

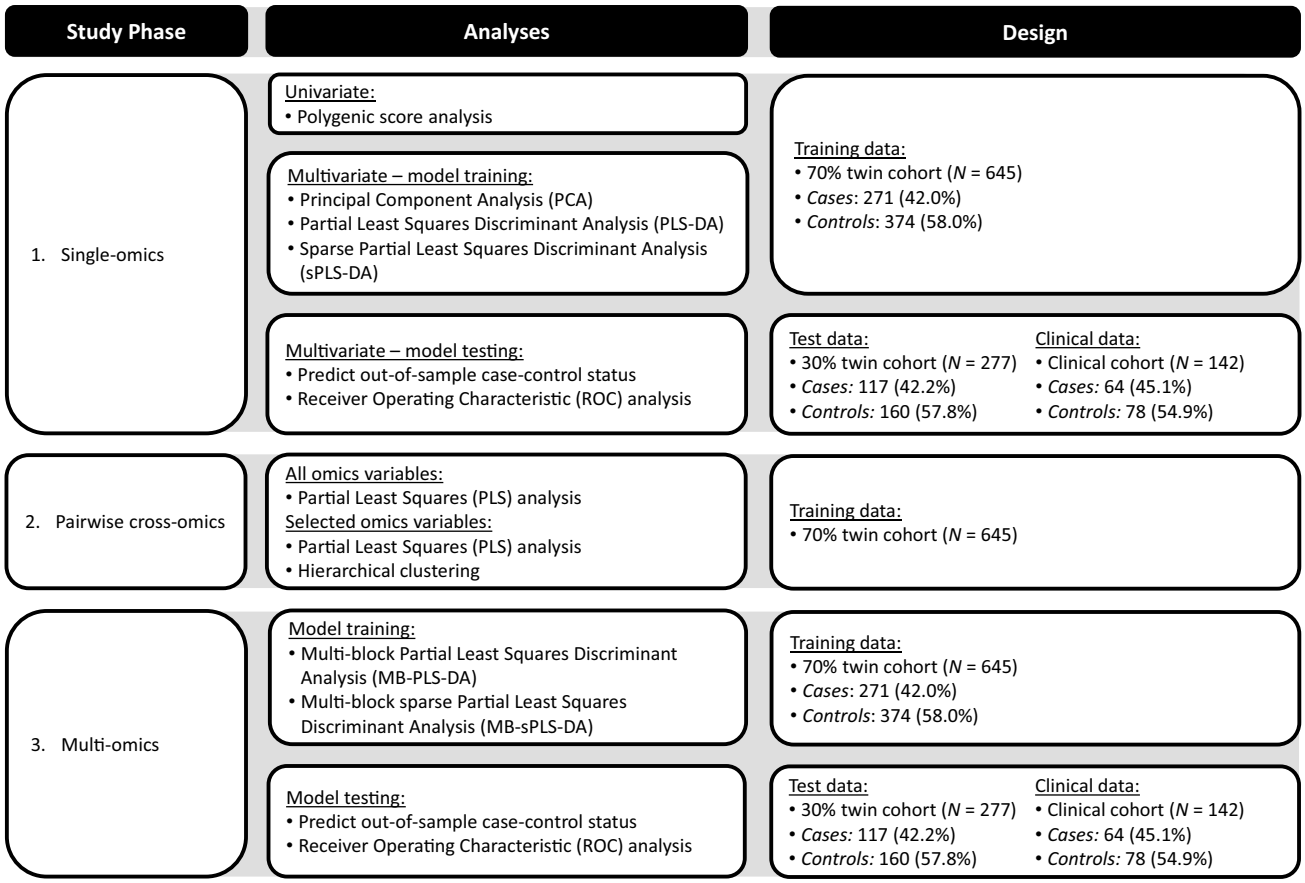


Fig. 1 Overview of the biomarker identification approach for childhood aggression—details of statistical analyses and data included in each analysis. We employed an analytical design comprising three phases: (1) single-omics analyses; (2) pairwise cross-omics analyses; and (3) multi-omics analyses. First, 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 "Materials and methods" section

Materials and Methods

Study Population and Procedures

Participants included 1494 twins (747 complete pairs) from the Netherlands Twin Register (Ligthart et al. 2019), and 189 children referred to the LUMC-Curium youth psychiatry clinic in the Netherlands that took part in the ACTION Biomarker Study (Aggression in Children: Unraveling gene-environment interplay to inform Treatment and InterventiON strategies) (Boomsma 2015; Bartels et al. 2018; Hagenbeek et al. 2020). Both cohorts collected first-morning urine samples and buccal-cell swabs with standardized protocols (for details, see Online Appendix A). In the twin cohort, we also collected buccal-cell swabs from parents and siblings of the twins. The current study included participants if they had aggression status, complete omics data for all three omics layers, and all relevant covariates (Table S1), as the current software does not allow for missing data. Parents provided written informed consent for their children and twin parents provided written informed consent for their own participation. Study approval was obtained from the Central Ethics Committee on Research Involving Human Subjects of the VU University Medical Center, Amsterdam (NTR 25th of May 2007 and ACTION 2013/41 and 2014.252), an Institutional Review Board certified by the U.S. Office of Human Research Protections (IRB number IRB00002991 under Federal-wide Assurance- FWA00017598; IRB/institute codes), and the Medical Ethical Committee of Leiden University Medical Center (B17.031, B17.032 and B17.040).

Aggressive Behavior

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

Omics Measurements

Detailed information on omics measurements is included in Online 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 3334 participants, including 1702 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.

Statistical Analyses

To define multi-omics biomarker panels capable of discriminating between cases and controls and explore the connections among the omics traits included in these panels, we employed an analytical design comprising three phases: (1) single-omics analyses; (2) pairwise cross-omics analyses; and (3) multi-omics analyses (Fig. 1). To avoid overfitting of the single- and multi-omics models, we randomly split the twin sample at the twin pair level into two subsets: 70% of the data for model training (training data), and 30% of the data for model testing (test data; Table 2). The clinical validity of the final single- and multi-omics models was evaluated in the clinical cohort (Table 2). We carried analyses out in the mixOmics R package (version 6.12.1) implemented in the R programming language 2020 (version 4.0.2) (R Core Team; Rohart et al. 2017).

Phase 1: Single-Omics Analyses

Univariate Polygenic Score Analyses In the training data, we first associated the transmitted- and non-transmitted PGSs with childhood aggression through generalized estimating equation (GEE) models. GEE models tested the association of each transmitted and non-transmitted PGS separately on the continuous mother-rated sum scores of the ASEBA CBCL Aggressive Behavior syndrome scale as assessed at the time of biological sample collection. All models included sex, age, genotype array, and the first 10 genetic principal components as covariates. We corrected for the correlation structure within families by using the "exchangeable" correlation structure, obtaining robust variance estimators (Rogers and Stoner 2016). A False Discovery Rate (FDR) of 5% for 45 PGSs was used to correct for multiple testing (p.adjust function in R), setting the significance threshold to

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

Trait [a]	N discovery GWA	Reference	Source original summary statistics
Childhood aggression	151,741	Ip et al 2021	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	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

[a]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"

$q \leq 0.05$, i.e., 5% of the significant results will be false positives (Benjamini and Hochberg 1995).

Multivariate Single-Omics Models To get a first insight into the dimensionality of the metabolomics data (90 variables), DNA methylation data (78,772 variables), and the PGSs (45 variables), we ran Principal Component Analysis (PCA) within each omics block in the training data (Table S2). To assess the ability of each of the three omics layers to correctly classify aggression status, we applied

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[a]	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)

MZ monozygotic

[a]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

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

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

The ability of the final single-omics models to predict out-of-sample case–control status was evaluated in the test and clinical datasets (*predict* function). For each new observation in the test and clinical datasets, this function calculates the predicted class (case/control) by estimating their predicted dummy variable (of the case–control status) using the maximum, Mahalanobis, or Centroids (Euclidian) distance (see Rohart et al. 2017). When using the maximum distance, the predicted class of a new observation is the class for which we observed the largest predicted dummy value. Both the Mahalanobis and centroids distances are centroid-based distances that predict the class of a new observation, so that the distance between its centroid and predicted scores is minimal. To predict out-of-sample case–control status, we used the best performing prediction distance, as was determined during model training (see Table S3).

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

Phase 2: Pairwise Cross-Omics Analyses

To highlight pairwise cross-omics relationships (i.e., DNA methylation-metabolomics, PGSs- metabolomics, and PGSs-DNA methylation), Partial Least Squares (PLS) regression models were constructed in canonical mode in the training data. Similar to canonical correlation analysis, canonical PLS regression aims to find linear combinations of the variables (canonical variates) to reduce the dimensionality of the data while maximizing the covariance between the variates (Rohart et al. 2017). CV has not yet been implemented for PLS in canonical mode. Consequently, we kept the smallest number of components as kept in either of the respective single-omics models (Table S3). Therefore, we included 3 components for the DNA methylation-metabolomics model, and 2 components for the PGS-metabolomics and PGS-DNA methylation models.

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

Phase 3: Multi-Omics Analyses

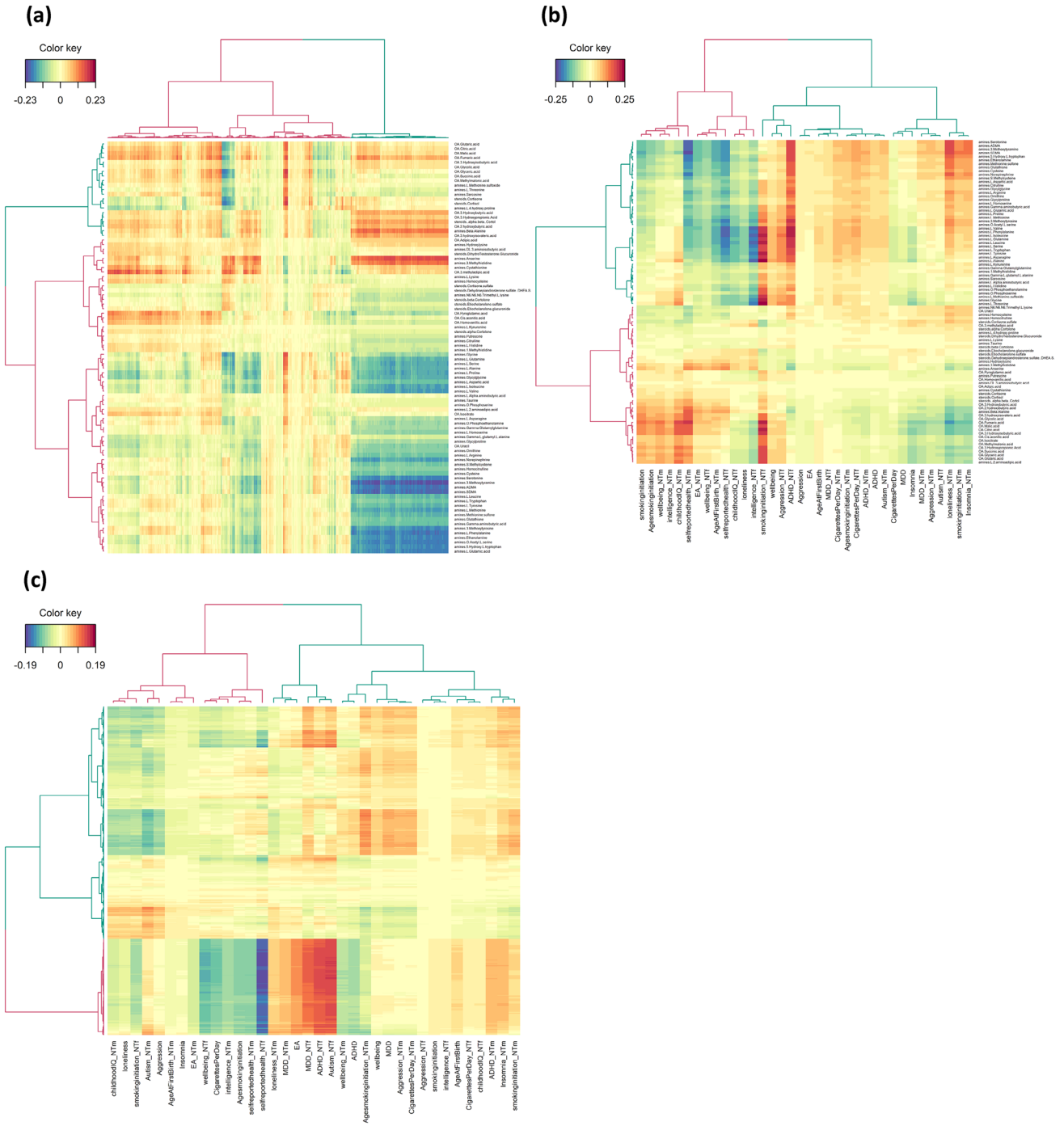
The multi-omics analysis was conducted through Data Integration Analysis for Biomarker discovery using Latent cOmponents (DIABLO) in the training data. DIABLO extends PLS-DA to multi-block PLS-DA (MB-PLS-DA), that aims at identifying correlated variables from multiple omics blocks that maximize the sample classification (Singh et al. 2019). The method requires a user-defined 'design matrix', that specifies the expected correlations among the omics blocks. The symmetric design matrix has the number of rows and columns equal to the number of omics blocks (i.e., 3), and contains values between 0 and 1. A 'full' design matrix denotes strong positive correlations among the omics blocks and sets the values among omics blocks close to or equal to one. A 'null' design matrix denotes weak or no correlations among omics blocks by setting values close to or equal to zero. The full design matrix optimizes correlations among the omics blocks, while the null design matrix optimizes the discrimination between samples (Rohart et al. 2017; Singh et al. 2019; Duruflé et al. 2020). We can also specify a design matrix with the empirical correlations among the omics blocks.

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

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

Biological Characterization

To facilitate biological interpretation, we describe the connections of the PGSs, CpGs, and metabolites, that were selected by the single-omics sPLS-DA models and those selected by the multi-omics MB-sPLS-DA models. For the multi-omics MB-sPLS-DA models we also identified correlation patterns that included high absolute correlations ($|r| \geq 0.60$) between omics traits of at least two omics blocks. As for the pairwise cross-omics models, these correlations comprise the correlations among the PLS variates. To construct the correlational patterns we denoted all instances where multiple omics variables were connected through bi- or trivariate correlations as a pattern. An example of a pattern comprising one variable of each omics block: high correlation of a PGS with a CpG and a high correlation of this CpG with a metabolites.



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

◂**Fig. 2** Clustered heatmaps of the relationships obtained by the pairwise cross-omics Partial Least Squares (PLS) regression models including only the omics variables as selected by the single-omics sparse Partial Least Squares Discriminant Analyses (sPLS-DA). We generated the hierarchical clustering using the Ward linkage algorithm on Euclidean distances of the PLS variates. For each dendrogram we identified two clusters (cluster 1 = pink, cluster 2 = blue). We have depicted positive relationships among the omics variables in red and negative relationships in blue. For the polygenic scores (PGSs), the '_NTm' suffix denotes non-transmitted maternal PGSs, the '_NTf' suffix denotes the non-transmitted paternal PGSs, and 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". For the metabolites, the 'amines.' prefix shows we measured these metabolites on the Liquid Chromatography Mass Spectrometry (LC–MS) amines platform, the 'steroids.' prefix shows we measured these metabolites on the LC–MS steroids platform, and the 'OA.' prefix shows we measured these metabolites on the Gas Chromatography (GC-) MS organic acids platform. **a** Relationships of the 1,614 CpGs and 90 metabolites included in the 3-component DNA methylation-metabolomics PLS model, where the selected CpGs are represented in the columns and the metabolomics traits in the rows. We included the cluster assignments and the full data matrix in Data S4-S5. **b** Relationships of the 36 PGSs and 90 metabolites included in the 2-component PGSs-metabolomics PLS model, where the selected PGSs are represented in the columns and the metabolomics traits in the rows. We included the cluster assignments and the full data matrix in Table S9 and Data S6, respectively. **c** Relationships of the 36 PGSs and 1,614 CpGs included in the 2-component PGSs-DNA methylation PLS model, where the selected PGSs are represented in the columns and the CpGs in the rows. We included the cluster assignments and the full data matrix in Data S7-S8 (Color figure online)

Results

Polygenic Prediction

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

Single-Omics Models for Childhood Aggression

We built cross-validated single-omics biomarker panels for childhood aggressive behavior based on sPLS-DA models including PGSs, DNA methylation, or metabolomics data. The final models comprised 11 transmitted and 25 non-transmitted PGSs, 1,614 CpGs, and all 90 metabolites (Table S3; Data S2; Fig. S8). The PGSs selected in the sPLS-DA model comprised the transmitted PGSs for aggression, ADHD, age at first birth, age at smoking initiation, number of cigarettes per day, educational attainment (EA), insomnia, loneliness, Major Depressive Disorder (MDD), smoking initiation, and wellbeing. The non-transmitted maternal PGSs were selected for aggression, ADHD, age at first birth, age at smoking initiation, Autism Spectrum Disorder (ASD), childhood IQ, number of cigarettes per day, EA, insomnia, intelligence, loneliness, MDD, self-reported health, smoking initiation, and wellbeing, and the non-transmitted paternal PGSs were selected for aggression, ADHD, ASD, childhood IQ, number of cigarettes per day, intelligence, MDD, self-reported health, smoking initiation, and wellbeing (Data S2). Trait enrichment analyses for all 1,614 selected CpGs in the sPLS-DA model showed the strongest enrichment for glucocorticoid exposure (i.e., administration of corticosteroid medication (Braun et al. 2019); OR = 18.02, $p = 5.34 \times 10^{-158}$), and household socioeconomic status in childhood (OR = 9.88, $p = 1.18 \times 10^{-13}$; Table S6). All three single-omics models showed low classification accuracy in the test and clinical data (Table S7; Fig. S9- S10).

Pairwise Cross-Omics Models

DNA Methylation-Metabolomics

In the pairwise DNA methylation-metabolomics model, we used hierarchical clustering and found two clusters of CpGs and of metabolites (Fig. 2a; Data S4). The DNA methylation cluster 1 contains 1151 (71.3%) of the CpGs selected by the sPLS-DA models for childhood aggression, and showed the strongest trait enrichments for household socioeconomic status in childhood (OR = 12.48, $p = 4.20 \times 10^{-14}$; Table S8). The 463 (28.7%) CpGs included in cluster 2 showed the strongest trait enrichment for glucocorticoid exposure (OR = 51.25, $p \leq 1.00 \times 10^{-308}$; Table S8). Metabolite cluster 1 contains 69 (76.7%) of the metabolites, including 55 amines, 7 organic acids and 7 steroids, while metabolite cluster 2 contains 21 metabolites (23.3%), including 5 amines, 13 organic acids, and 3 steroids. The average correlation between the DNA methylation and metabolomics blocks, i.e., the correlation among all PLS variates of all components simultaneously, was 0.18 ($q = 6.07 \times 10^{-15}$). We observed the highest absolute correlations between metabolites of cluster 1 and CpGs included in cluster 2 (Fig. 2a; Data S4-S5). Specifically, the amines 3-methoxytyramine (r M = − 0.21, r range: − 0.23–0.20), asymmetric dimethylarginine (ADMA, r M = − 0.20, r range: − 0.21–− 0.20), L-glutamic acid (r = − 0.20), L-phenylalanine (r M = − 0.20, r range: − 0.21–− 0.20), O-acetyl-L-serine (r = − 0.20), and symmetric dimethylarginine (SDMA, r M = − 0.21, r range: − 0.22–− 0.20) show negative correlations with these CpGs, indicating that increased levels of these urinary metabolites associated with hypomethylation at cluster 2 CpG sites.

PGSs-Metabolomics

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

PGSs-DNA Methylation

For the pairwise PGSs-DNA methylation model, we again applied hierarchical clustering to find two clusters for the PGSs and DNA methylation data (Fig. 2c; Data S7). PGS cluster 1 contains 22 of the PGSs (61.1%) selected by the sPLS-DA model for childhood aggression. This cluster contained 6 transmitted PGSs, 9 non-transmitted maternal PGSs, and 7 non-transmitted paternal PGSs. The 14 PGSs (38.9%) included in cluster 2 comprised 5 transmitted PGSs, 6 non-transmitted maternal PGSs, and 3 non-transmitted paternal PGSs. The DNA methylation cluster 1 contained 1142 (70.8%) of the CpGs selected by the sPLS-DA models for childhood aggression, and showed the strongest trait enrichments for household socioeconomic status in childhood (OR = 12.58, $p = 3.68 \times 10^{-14}$; Table S10). The 472 (29.2%) CpGs included in cluster 2 showed the strongest trait enrichment for glucocorticoid exposure (OR = 50.00, $p \leq 1.00 \times 10^{-}308$; Table S10). The average correlation between the PGSs and DNA methylation block, i.e., the correlation among all PLS variates of all components simultaneously, was 0.28 ($q = 9.57 \times 10^{-24}$). The most negative correlations were observed between the non-transmitted paternal self-reported health PGS with CpGs of cluster 2 (r M = − 0.18, r range: − 0.19–− 0.15), and the highest positive correlations between the cluster 2 CpGs and the non-transmitted paternal ASD PGS (r M = 0.151, r range: 0.150–0.154; Fig. 2c; Data S8). This indicates that characteristics tagged by ASD and self-reported health correlated to paternal construction of environments influencing offspring buccal DNA methylation levels.

Multi-Omics Model for Childhood Aggression

We built multi-omics panels for childhood aggressive behavior based on multi-block sPLS-DA (MB-sPLS-DA) models, including PGSs, DNA methylation, and metabolomics data. Here, we report the multi-omics model with an empirical design matrix and results for the null and full design matrices can be found in Online Appendices C and D, respectively. After cross-validation, the optimal 5-component model included 14 transmitted and 30 non-transmitted PGSs, 746 CpGs, and all 90 metabolites (Table S5; Data S9; Fig. S11). The multi-omics model selected all 30 non-transmitted PGS, and except for the transmitted PGS for childhood IQ, also all the transmitted PGSs (Data S9). Out of the 746 CpGs selected by the multi-omics model 204 (27.3%) were also selected by the single-omics DNA methylation model (Data S2; Data S9). Trait enrichment analyses for all selected CpGs in the MB-sPLS-DA model showed the strongest enrichment for gender (OR = 3.90, $p = 3.97 \times 10^{-31}$), and breast cancer risk (OR = 342.60, $p = 1.44 \times 10^{-21}$; Table S11). Multi-omics prediction of aggression case–control status in the test data showed an improvement in the prediction (BER = 0.47–0.52; Table S12; Fig. S12) as compared to single-omics models including only the PGSs or metabolomics data, but not as compared to the model including only the DNA methylation data (Table S7). In contrast, in the clinical cohort, the average classification accuracy was poorer in the multi-omics model (BER = 0.53–0.57; Table S12) than for the single-omics models (Table S7; Fig. S13).

The average correlations between each omics block in the multi-omics model, i.e., the correlation among all PLS variates of all components simultaneously, were $r = 0.19$ ($q = 2.13 \times 10^{-27}$) for PGSs-DNA methylation, $r = 0.13$ ($q = 1.78 \times 10^{-12}$) for PGSs-metabolomics, and $r = 0.15$ ($q = 3.76 \times 10^{-16}$) for DNA methylation-metabolomics. We observed high absolute correlations ($|r| \geq 0.60$) between 2 selected PGSs, 14 CpGs, and 13

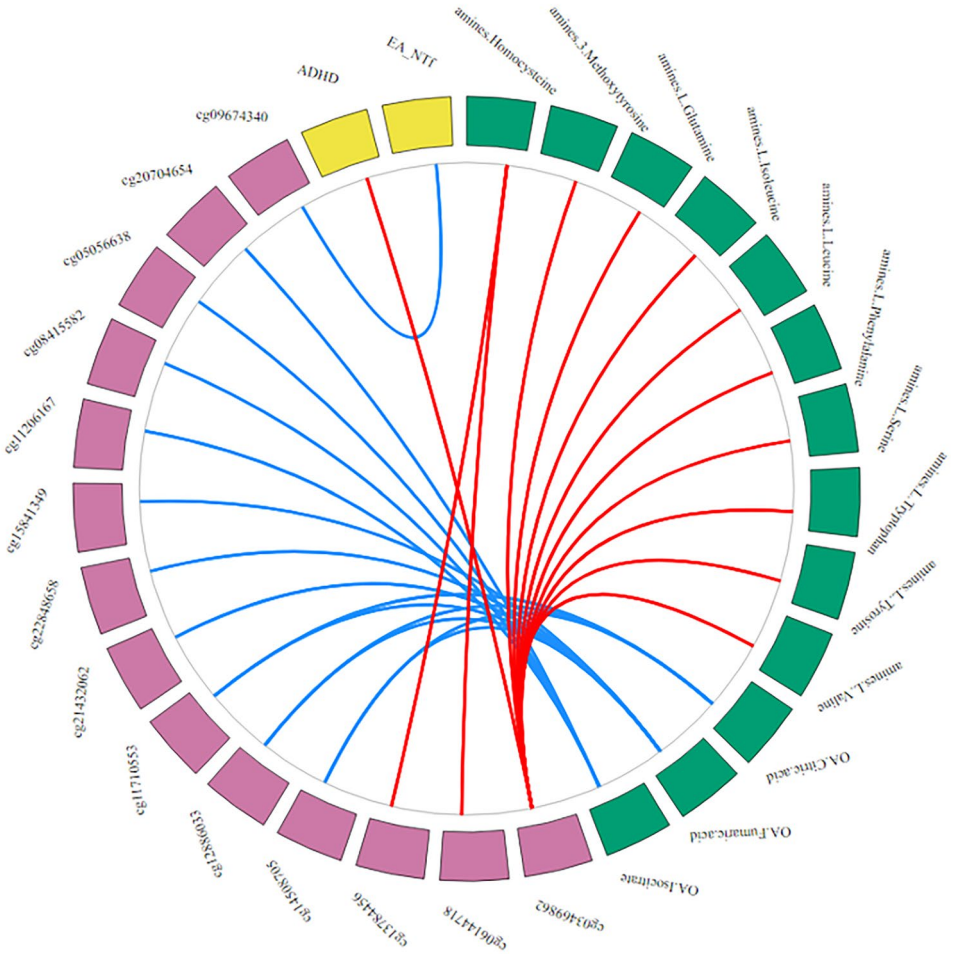


Fig. 3 Strong cross-omics connections of the multi-omics traits identified in the 5-component multi-block sparse Partial Least Squares Discriminant Analysis (MB-sPLS-DA) model including the empirical design matrix. The outer ring depicts the PGSs, CpGs, and metabolites in yellow, pink, and green, respectively. 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 (Color figure online)

metabolite, that we summarize in five sets of correlational patterns (Fig. 3; Data S10; Table S13). Correlation pattern 1 comprises high negative correlations of citric and fumaric acid with cg12886033 (chr1:65,449,013), cg14508705 (chr1:172,360,182), and cg11710553 (chr4:105,892,960), and high negative correlations of fumaric acid with cg21432062 (chr3:4,908,643), cg22848658 (chr6:135,354,586), and cg15841349 (chr12:129,348,564). The second correlational pattern is characterized by high negative correlations between isocitrate with cg05056638 (chr8:24,800,824), cg08415582 (chr8:57,030,523), cg11206167 (chr5:42,924,367), and cg20704654 (chr20:30,072,118). Correlation pattern 3 contains the high positive correlation of homocysteine with cg13784456 (chr10:132,970,405) and cg06144718 (chr10:133,048,392). The fourth correlational pattern includes high positive correlations between cg03469862 (chr11:68,924,853) with the transmitted PGS for ADHD, 3-methoxytyrosin, L-isoleucine, L-leucine, L-phenylalanine, L-tryptophan, L-valine, L-glutamine,

L-tyrosine, and L-serine. The final correlation pattern comprises a high negative correlation between the non-transmitted by father PGS for EA and cg09674340 (chr1:202,509,286).

Discussion

This study comprised an integrative multi-omics analysis of childhood aggressive behavior to identify a multi-omics biomarker panel and to investigate the correlations among the omics blocks included in this panel. In our training data comprising 645 twins (cases = 42.0%, controls = 58.0%) we applied multivariate statistical methods to analyze and integrate transmitted and paternal and maternal non-transmitted PGSs for childhood aggression and for 14 traits genetically correlated with aggression (45 PGSs total), 78,772 CpGs, and 90 metabolites (Fig. 1). We build single-omics biomarker panels for each of the omics blocks, that selected 31 PGSs, 1614 CpGs, and 90 metabolites to discriminate between aggression cases and controls. The markers selected in single-omics models had poor predictive performance in our test ($N = 277$, cases = 42.2%, controls = 57.8%) and clinical datasets ($N = 142$, cases = 45.1%, controls = 54.9%). We explored the pairwise correlations of the omics variables selected by the single-omics models and found that indirect genetic effects for ADHD, ASD, intelligence, smoking initiation, and self-reported health, connected most strongly with buccal DNA methylation and urinary amino acid levels in children and that higher amino acid levels associate with hypomethylation in a cluster of CpG sites.

The cross-omics analyses generally showed stronger connections among non-transmitted polygenic scores and DNA methylation or metabolite levels. We consider this a noteworthy finding which may suggest that parents shape environments that impact on offspring DNA methylation and metabolites and that these processes can be detected when adding non-transmitted polygenic scores in a multi-omics approach. The stronger connections for non-transmitted PGSs may thus imply that the CpGs selected by our model not merely capture environmental effects but are more sensitive to genetic nurturing on aggression-related traits. For future understanding of 'environmental' mechanisms, studies of aggression and other childhood phenotypes and disorders would benefit from an exploration of the associations of large numbers of non-transmitted PGSs, based on multiple discovery GWASs and CpG sites.

The multi-omics panel selected 44 PGSs, 746 CpGs, and 90 metabolites, which also had poor predictive performance in our test and clinical datasets. We described five sets of correlational patterns with high absolute correlations ($|r| \geq 0.60$) of aggression-related omics variables selected by the multi-omics model. Correlation patterns 1 and 3 likely capture the DNA methylation process itself, as homocysteine is involved in the methionine cycle that transfers the methionine methyl group to DNA methyltransferase (Selhub 1999), and tricarboxylic acid (TCA) cycle metabolites, such as citric and fumaric acid, play a role in histone acetylation, and histone and DNA demethylation (Martínez-Reyes and Chandel 2020). The CpGs and their connected metabolites included in patterns 1, 2, and 4 were previously associated with numerous traits, including ageing, inflammatory bowel disease (IBD), prostate cancer, and Myelodysplastic Syndrome (Ooi et al. 2011; Schicho et al. 2012; Dawiskiba et al. 2014; Rist et al. 2017; Lima et al. 2021; Yuan et al. 2021; Xiong et al. 2022). Pattern 5 links a smoking and low birth weight-associated CpG to indirect genetic effects for EA. The negative association of a smoking-associated CpG with indirect genetic effects for EA may indicate that rearing environment influences smoking-related DNA methylation. Pattern 5 may be regarded as being in line with a previous study that reported significant negative correlations of indirect parental EA effect with offspring smoking initiation, number of cigarettes per day, and smoking cessation (Wu et al. 2021), and the well-established phenotypic association of low paternal educational attainment with low offspring birth weight (e.g., Meng and Groth 2018). Overall, the multi-omics correlational patterns associated with a range of traits that link aggression-related omics variables to biological processes related to inflammation, carcinogens, ageing, sex differentiation, intelligence, and smoking.

We explored how sensitive our results are to different specifications of the design matrices and found that these resulted in similar predictive abilities, but selected different numbers of only partially overlapping PGSs and CpGs (Online Appendices C and D). In line with expectations, the predictive ability of the null design matrix was slightly better than for the empirical design matrix, reflecting that such a model focuses on selecting discriminatory variables (Singh et al. 2019; Duruflé et al. 2020). We expected that the model with the full design model sacrificed predictive accuracy to select discriminatory variables that are also highly correlated, but this model had the lowest classification error rate as compared to the models with an empirical or null design matrix in the test and clinical data. We note that, regardless of the design matrix, our predictive accuracy was poor, which included observed Balanced Error Rates (BERs) below chance level (≤ 0.50). This indicates that the identified biological signals are not strongly predictive of childhood aggression. Nevertheless, as GWAS have clearly demonstrated, even very small effects may generate novel biological insights. Here, we show that multi-omics models with differently specified design matrices differed not only in how many and which PGSs and CpGs were selected, but each multi-omics model provided unique insight into the

correlations among the omics variables selected for their association with childhood aggression. In the current study, we relied on the cross-validation results to select the number of variables and components in the model. However, to aid in biological interpretation, other choices can be made.

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

The CpGs in the single- or multi-omics models did not overlap with the top differentially methylated CpGs for physical aggression as observed in buccal-cells (Cecil et al. 2018b), and were not significantly associated with aggression in a recent blood-based EWAS meta-analysis (van Dongen et al. 2021). That our models selected no CpGs overlapping with these studies was unsurprising, since it was previously shown that the top CpGs from these studies were not associated with aggressive behavior in the buccal DNA methylation data in the NTR and LUMC-Curium cohorts (van Dongen et al. 2021). Trait enrichment of the CpGs selected by the single- and multi-omics models reported enrichment of known aggression risk factors, such as socioeconomic status (Miller and Tolan 2019; Bellair et al. 2019; Hendriks et al. 2020), childhood malnutrition (Liu 2004; Vaughn et al. 2016), and pre- and perinatal risk factors (Van Adrichem et al. 2020). Moreover, we observed trait enrichment for several syndromes, e.g., Down syndrome, that are characterized by intellectual disability, developmental delay, and sometimes by aggressive behaviors among affected individuals. Several of these syndromes, including Klinefelter and Claes-Jensen syndrome, are linked to genetic abnormalities of the sex chromosomes, and CpGs in the single- and mult-omics models are enriched for steroid hormone-related traits, such as glucocorticoid exposure. Together, these results suggest that the CpGs in the single- and multi-omics models reflect the well-established gender differences in aggressive behavior. We note that trait enrichment analyses might be less sensitive to EPIC array-specific CpG sites as less than 20% of the current version of the EWAS atlas is comprised of EPIC studies (October 2022; Xiong et al. 2022). In addition to the enrichment of glucocorticoid exposure in CpGs selected by the single-omics model, we observed a high correlation between cortisol and cg05153029 (chr20:19,769,815, $r = 0.61$), that associates with glucocorticoid exposure in the EWAS atlas (full design matrix MB-sPLS-DA model). Cortisol and other hypothalamus–pituitary–adrenal (HPA)-axis molecules have been implicated in aggressive behavior, suggesting a role of the stress response system in aggressive behavior (Hagenbeek et al. 2016). Epigenetic programming of the HPA-axis is influenced by pre- and perinatal factors, such as maternal behavior as observed in rats (Weaver et al. 2004), and maternal stress and early life adversity (Mulligan et al. 2012; Hompes et al. 2013; Jiang et al. 2019), and lower average global DNA methylation levels were reported in patients with Cushing's Syndrome (CS) in remission, a model for long-standing excessive glucocorticoid (cortisol) exposure (Glad et al. 2017). Thus, epigenetic mechanisms may mediate the association between cortisol and childhood aggression, similarly as the DNA methylation mediates the association between cortisol stress reactivity and childhood trauma (Wrigglesworth et al. 2019). A highly promising approach to investigate whether epigenetic mechanisms mediate the association between steroid hormones and childhood aggression would be Mendelian Randomization (Sanderson 2021; Carter et al. 2021).

The single- and multi-omics models selected all 90 LC- and GC–MS metabolites for inclusion, which might be explained by the low expression differences between aggression cases and controls of the urinary amines and organic acids in this sample (Hagenbeek et al. 2020). Another plausible explanation regards the fact that we generated the metabolomics data on three targeted platforms, chosen because they cover relevant metabolites involved in neurotransmitter, inflammation, and steroid hormone pathways that associate with aggression (Hagenbeek et al. 2016). To date, no metabolomics platform can capture the entire metabolome. In addition, within a platform, technical challenges may cause compounds becoming undetectable or not quantifiable. This was the case for the steroid platform being less successful in quantifying (conjugated) sex hormones. Coverage could also be extended to include other metabolite classes, preferably in a hypothesis-free manner by employing non-targeted metabolomics platforms. The five sets of omics variables with high cross-omics correlations in the multi-omics model contained 13 metabolites. These metabolites comprised 8 amino acids, 3 metabolites involved in the TCA cycle, the dopaminergic trace amine 3-methoxytyrosine, and homocysteine, which is involved in cysteine and methionine metabolism. This is in line with previous studies reporting association of metabolites with aggressive behavior (Gulsun et al. 2016; Hagenbeek et al. 2016, 2020; Chen et al. 2020).

This is the first multi-omics study that includes DNA methylation profiles from buccal and metabolomics in urine. Most of the earlier large-scale omics studies were conducted in blood samples. By obtaining omics measurements in easily accessible peripheral tissues (urine and buccal-cells) we could obtain multi-omics data in a substantially larger (~tenfold) sample compared to most previous multi-omics studies for psychiatric traits, such as depression and suicide risk (Bhak et al. 2019) or post-traumatic stress disorder (Dean et al. 2019), that relied on small training samples (range: 126–165). In the current study, we corrected the omics data for sex, age, and technical covariates, which can be considered as a minimal set of covariates. Future studies may consider including other confounding factors, such as parental smoking, dietary factors, or body mass index. In evaluating the validity of the PGS and multi-omics models in the clinical cohort, the non-transmitted PGSs could not be assessed, as no parental genotypes were available in the clinical cohort. Additional validation in cohorts with complete omics data that applied the same metabolomics and DNA methylation platforms are of large interest, but currently do not exist. We corrected the omics data for sex and selected children for inclusion into the study before the onset of puberty (Hagenbeek et al., 2020). Ideally, however, future multi-omics studies in children would also allow for sex-stratified analyses as the predictive accuracy of our models might have been impacted by the inclusion of a smaller number of girls (~25%) in the clinical cohort compared to the twin cohort (~50%). Our design is optimal in nearly all other aspects. The twin and clinical cohorts collected all data, including biomarkers, at the same time, using the same arrays and platforms and with similar protocols, thus decreasing unwanted sources of heterogeneity that are often present in sequential multi-omics designs (Wörheide et al. 2021).

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

Conclusions

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

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10519-022-10126-7.

Acknowledgements 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.

Authors Contributions Conceptualization, FAH, JD, MB, SD and DIB; methodology, SD; formal analysis, FAH, RP and JJH; investigation, FAH, PJR, OFC, RRJMV, MB and DIB; resources, ACH, CK, EAE and TH; data curation, CEMB; writing—original draft preparation, FAH, JD, RP, MB and DIB; writing—review and editing, all authors; visualization, FAH; supervision, JD, MB and DIB; funding acquisition, ACH, CK, OFC, VF, TH, RRJMV, MB and DIB. All authors have read and agreed to the published version of the manuscript.

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

Data Availability The standardized protocol for large scale collection of urine and buccal-cell samples in the home situation as developed for the ACTION Biomarker Study in children available at http://www.action-euproject.eu/content/data-protocols. The data of the Netherlands Twin Register (NTR) ACTION Biomarker Study may be accessed, upon approval of the data access committee, through the NTR (https://tweelingenregister.vu.nl/information_for_researchers/working-with-ntr-data).

Code Availability Analysis code is available upon request from the corresponding author.

Declarations

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

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

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

Consent for publication Not applicable.

References

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

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

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

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

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

Bates TC, Maher BS, Medland SE et al (2018) The nature of nurture: using a virtual-parent design to test parenting effects on children's educational attainment in genotyped families. Twin Res Hum Genet 21:73–83. https://doi.org/10.1017/thg.2018.11

Beck JJ, Hottenga J-J, Mbarek H et al (2019) Genetic similarity assessment of twin-family populations by custom-designed genotyping array. Twin Res Hum Genet 22:210–219. https://doi.org/10.1017/thg.2019.41

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

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B 57:289–300

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

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

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

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

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

Carter AR, Sanderson E, Hammerton G et al (2021) Mendelian randomisation for mediation analysis: current methods and challenges for implementation. Eur J Epidemiol 36:465–478. https://doi.org/10.1007/s10654-021-00757-1

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

Cecil CAM, Walton E, Pingault JB et al (2018b) DRD4 methylation as a potential biomarker for physical aggression: an epigenome-wide, cross-tissue investigation. Am J Med Genet Part

B Neuropsychiatr Genet 177:746–764. https://doi.org/10.1002/ajmg.b.32689
Chen X, Xu J, Tang J et al (2020) Dysregulation of amino acids and lipids metabolism in schizophrenia with violence. BMC Psychiatry 20:1–11. https://doi.org/10.1186/s12888-020-02499-y
Dawiskiba T, Deja S, Mulak A et al (2014) Serum and urine metabolomic fngerprinting in diagnostics of inflammatory bowel diseases. World J Gastroenterol 20:163–174. https://doi.org/10.3748/wjg.v20.i1.163
Dean KR, Hammamieh R, Mellon SH et al (2019) Multi-omic biomarker identification and validation for diagnosing warzone-related post-traumatic stress disorder. Mol Psychiatry. https://doi.org/10.1038/s41380-019-0496-z
Demontis D, Walters RK, Martin J et al (2019) Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nat Genet 51:63–75. https://doi.org/10.1038/s41588-018-0269-7
Duruflé H, Selmani M, Ranocha P et al (2020) A powerful framework for an integrative study with heterogeneous omics data: from univariate statistics to multi-block analysis. Brief Bioinform. https://doi.org/10.1093/bib/bbaa166
Ehli EA, Abdellaoui A, Fedko IO et al (2017) A method to customize population-specific arrays for genome-wide association testing. Eur J Hum Genet 25:267–270. https://doi.org/10.1038/ejhg.2016.152
Fergusson DM, Horwood LJ, Ridder EM (2005) Show me the child at seven: the consequences of conduct problems in childhood for psychosocial functioning in adulthood. J Child Psychol Psychiatry Allied Discip 46:837–849. https://doi.org/10.1111/j.1469-7610.2004.00387.x
Galili T (2015) dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. Bioinformatics 31:3718–3720. https://doi.org/10.1093/bioinformatics/btv428
Glad CAM, Andersson-Assarsson JC, Berglund P et al (2017) Reduced DNA methylation and psychopathology following endogenous hypercortisolism—a genome-wide study. Sci Rep 7:1–11. 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. 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 across raters, instruments, and age. Transl Psychiatry 11:413. https://doi.org/10.1038/s41398-021-01480-x
Jakovljevic M, Jakovljevic I (2019) A Transdisciplinary integrative approach for precision psychiatry. In: Kim YK (ed) Frontiers in psychiatry. Springer, Singapore, pp 399–428
Jansen PR, Watanabe K, Stringer S et al (2019) Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. Nat Genet 51:394–403. https://doi.org/10.1038/s41588-018-0333-3
Jiang S, Postovit L, Cattaneo A et al (2019) Epigenetic modifications in stress response genes associated with childhood trauma. Front Psychiatry 10:1–19. https://doi.org/10.3389/fpsyt.2019.00808
Kassing F, Godwin J, Lochman JE, Coie JD (2019) Using early childhood behavior problems to predict adult convictions. J Abnorm Child Psychol 47:765–778. https://doi.org/10.1007/s10802-018-0478-7
Kong A, Thorleifsson G, Frigge ML et al (2018) The nature of nurture: effects of parental genotypes. Science 359:424–428. https://doi.org/10.1126/science.aan6877
Lê Cao KA, Boitard S, Besse P (2011) Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. BMC Bioinform. https://doi.org/10.1186/1471-2105-12-253
Lee JJ, Wedow R, Okbay A et al (2018) Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat Genet 50:1112–1121. https://doi.org/10.1038/s41588-018-0147-3
Ligthart L, van Beijsterveldt CEM, Kevenaar ST et al (2019) The Netherlands twin register: longitudinal research based on twin and twin-family designs. Twin Res Hum Genet 22:623–636. https://doi.org/10.1017/thg.2019.93
Lima AR, Pinto J, Amaro F et al (2021) Advances and perspectives in prostate cancer biomarker discovery in the last 5 years through tissue and urine metabolomics. Metabolites 11:181. https://doi.org/10.3390/metabo11030181
Liu J (2004) Malnutrition at age 3 years and externalizing behavior problems at ages 8, 11, and 17 years. Am J Psychiatry 161:2005–2013. https://doi.org/10.1176/appi.ajp.161.11.2005
Liu M, Jiang Y, Wedow R et al (2019) Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. Nat Genet 51:237–244
Martínez-Reyes I, Chandel NS (2020) Mitochondrial TCA cycle metabolites control physiology and disease. Nat Commun 11:1–11. https://doi.org/10.1038/s41467-019-13668-3
Masud H, Ahmad MS, Cho KW, Fakhr Z (2019) Parenting styles and aggression among young adolescents: a systematic review of literature. Community Ment Health J 55:1015–1030. https://doi.org/10.1007/s10597-019-00400-0
Meng Y, Groth SW (2018) Fathers count: the impact of paternal risk factors on birth outcomes. Matern Child Health J 22:401–408. https://doi.org/10.1007/s10995-017-2407-8
Miller GM, Tolan PH (2019) The influence of parenting practices and neighborhood characteristics on the development of childhood aggression. J Community Psychol 47:135–146. https://doi.org/10.1002/jcop.22105
Mitjans M, Seidel J, Begemann M et al (2019) Violent aggression predicted by multiple pre-adult environmental hits. Mol Psychiatry 24:1549–1564. https://doi.org/10.1038/s41380-018-0043-3
Moran S, Arribas C, Esteller M (2016) Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome

enriched in enhancer sequences. Epigenomics 8:389–399. https://doi.org/10.2217/epi.15.114
Mulligan CJ, D'Errico NC, Stees J, Hughes DA (2012) Methylation changes at NR3C1 in newborns associate with maternal prenatal stress exposure and newborn birth weight. Epigenetics 7:853–857. https://doi.org/10.4161/epi.21180
Odintsova VV, Roetman PJ, Ip HF et al (2019) Genomics of human aggression. Psychiatr Genet 29:170–190. https://doi.org/10.1097/YPG.0000000000000239
Ooi M, Nishiumi S, Yoshie T et al (2011) GC/MS-based profiling of amino acids and TCA cycle-related molecules in ulcerative colitis. Inflamm Res 60:831–840. https://doi.org/10.1007/s00011-011-0340-7
Pinu FR, Beale DJ, Paten AM et al (2019) Systems biology and multi-omics integration: viewpoints from the metabolomics research community. Metabolites 9:1–31. https://doi.org/10.3390/metabo9040076
R Core Team: A language and environment for statistical computing. https://www.r-project.org/
Radwan K, Coccaro EF (2020) Comorbidity of disruptive behavior disorders and intermittent explosive disorder. Child Adolesc Psychiatry Ment Health 14:24. https://doi.org/10.1186/s13034-020-00330-w
Rist MJ, Roth A, Frommherz L et al (2017) Metabolite patterns predicting sex and age in participants of the karlsruhe metabolomics and nutrition (KarMeN) study. PLoS ONE 12:1–21. https://doi.org/10.1371/journal.pone.0183228
Rogers P, Stoner J (2016) Modification of the sandwich estimator in generalized estimating equations with correlated binary outcomes in rare event and small sample settings. Am J Appl Math Stat 3:243–251
Rohart F, Gautier B, Singh A, Lê Cao KA (2017) mixOmics: an R package for 'omics feature selection and multiple data integration. PLoS Comput Biol 13:1–20. https://doi.org/10.1371/journal.pcbi.1005752
Sanderson E (2021) Multivariable mendelian randomization and mediation. Cold Spring Harb Perspect Med 11:a038984. https://doi.org/10.1101/cshperspect.a038984
Savage JE, Jansen PR, Stringer S et al (2018) Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. Nat Genet 50:912–919. https://doi.org/10.1038/s41588-018-0152-6
Schicho R, Shaykhutdinov R, Ngo J et al (2012) Quantitative metabolomic profiling of serum, plasma, and urine by 1H NMR spectroscopy discriminates between patients with inflammatory bowel disease and healthy individuals. J Proteome Res 11:3344–3357. https://doi.org/10.1021/pr300139q
Selhub J (1999) Homocystein metabolism. Annu Rev Nutr 19:217–246. https://doi.org/10.1146/annurev.nutr.19.1.217
Siever LJ (2008) Neurobiology of aggression and violence. Am J Psychiatry 165:429–442
Singh A, Shannon CP, Gautier B et al (2019) DIABLO: an integrative approach for identifying key molecular drivers from multi-omics assays. Bioinformatics 35:3055–3062. https://doi.org/10.1093/bioinformatics/bty1054
Sinke L, van Iterson M, Cats D, et al (2019) DNAmArray: Streamlined workflow for the quality control, normalization, and analysis of Illumina methylation array data (2.1). Zenodo. https://molepi.github.io/DNAmArray_workflow/
Subramanian I, Verma S, Kumar S et al (2020) Multi-omics data integration, interpretation, and its application. Bioinform Biol Insights. https://doi.org/10.1177/1177932219899051
Tibshirani R (1996) Regression shrinkage and selection via the lasso. J R Stat Soc Ser B (methodol) 58:267–288
Van Adrichem DS, Huijbregts SCJ, Van Der Heijden KB et al (2020) Aggressive behavior during toddlerhood: interrelated effects of prenatal risk factors, negative affect, and cognition. Child Neuropsychol 26:982–1004. https://doi.org/10.1080/09297049.2020.1769582
van der Laan CM, Morosoli-García JJ, van de Weijer SGA et al (2021) Continuity of genetic risk for aggressive behavior across the life-course. Behav Genet. https://doi.org/10.1007/s10519-021-10076-6
van Dongen J, Hagenbeek FA, Suderman M et al (2021) DNA methylation signatures of aggression and closely related constructs: a meta-analysis of epigenome-wide studies across the lifespan. Mol Psychiatry. https://doi.org/10.1038/s41380-020-00987-x
Vaughn M, Salas-Wright C, Naeger S et al (2016) Childhood reports of food neglect and impulse control problems and violence in adulthood. Int J Environ Res Public Health 13:389. https://doi.org/10.3390/ijerph13040389
Vuoksimaa E, Rose RJ, Pulkkinen L et al (2020) Higher aggression is related to poorer academic performance in compulsory education. J Child Psychol Psychiatry. https://doi.org/10.1111/jcpp.13273
Watanabe K, Stringer S, Frei O et al (2019) A global overview of pleiotropy and genetic architecture in complex traits. Nat Genet 51:1339–1348. https://doi.org/10.1038/s41588-019-0481-0
Weaver ICG, Cervoni N, Champagne FA et al (2004) Epigenetic programming by maternal behavior. Nat Neurosci 7:847–854. https://doi.org/10.1038/nn1276
Whipp AM, Korhonen T, Raevuori A et al (2019) Early adolescent aggression predicts antisocial personality disorder in young adults: a population-based study. Eur Child Adolesc Psychiatry 28:341–350. https://doi.org/10.1007/s00787-018-1198-9
Whipp AM, Vuoksimaa E, Bolhuis K et al (2021a) Teacher-rated aggression and co-occurring behaviors and emotional problems among schoolchildren in four population-based European cohorts. PLoS ONE 16:e0238667. https://doi.org/10.1371/journal.pone.0238667
Whipp AM, Vuoksimaa E, Korhonen T et al (2021b) Ketone body 3-hydroxybutyrate as a biomarker of aggression. Sci Rep 11:5813. https://doi.org/10.1038/s41598-021-84635-6
Wörheide MA, Krumsiek J, Kastenmüller G, Arnold M (2021) Multi-omics integration in biomedical research—a metabolomics-centric review. Anal Chim Acta 1141:144–162. https://doi.org/10.1016/j.aca.2020.10.038
Wray NR, Ripke S, Mattheisen M et al (2018) Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nat Genet 50:668–681. https://doi.org/10.1038/s41588-018-0090-3
Wrigglesworth J, Ancelin ML, Ritchie K, Ryan J (2019) Association between DNA methylation of the KITLG gene and cortisol levels under stress: a replication study. Stress 22:162–168. https://doi.org/10.1080/10253890.2018.1519019
Wu Y, Zhong X, Lin Y et al (2021) Estimating genetic nurture with summary statistics of multigenerational genome-wide association studies. Proc Natl Acad Sci 118:e2023184118. https://doi.org/10.1073/pnas.2023184118
Xiong Z, Yang F, Li M et al (2022) EWAS open platform: integrated data, knowledge and toolkit for epigenome-wide association study. Nucleic Acids Res 50:D1004–D1009. https://doi.org/10.1093/nar/gkab972
Yuan Y, Zhao J, Li T et al (2021) Integrative metabolic profile of myelodysplastic syndrome based on UHPLC–MS. Biomed Chromatogr. https://doi.org/10.1002/bmc.5136