

# Genetic and epigenetic correlates of co-occurring behavioral problems in early childhood

**Yunpeng Wang**

`yunpeng.wang@psykologi.uio.no`

University of Oslo <https://orcid.org/0000-0001-9831-1090>

**Mona Bekkhus**

PROMENTA Research Centre, Department of Psychology, University of Oslo

**Stella Tsotsi**

**Rosa Mulder**

Erasmus MC <https://orcid.org/0000-0002-2382-1704>

**Zoe Hart**

**Christian Page**

<https://orcid.org/0000-0002-1897-3666>

**Nikolai Olavi Czajkowski**

Norwegian Institute of Public Health

**Wanlin Liu**

**Espen Røysamb**

**Ragnhild Nes**

**Aurora Oftedal**

**Robert Lyle**

Department of Medical Genetics, Oslo University Hospital <https://orcid.org/0000-0001-6317-732X>

**Laurie Hannigan**

**Ole Andreassen**

Oslo University Hospital & Institute of Clinical Medicine, University of Oslo <https://orcid.org/0000-0002-4461-3568>

**Vilte Baltramonaityte**

University of Bath <https://orcid.org/0000-0002-9776-735X>

**Esther Walton**

<https://orcid.org/0000-0002-0935-2200>

**Charlotte Cecil**

, Erasmus Medical Centre, University Medical Center Rotterdam <https://orcid.org/0000-0002-2389-5922>

## Article

### Keywords:

**Posted Date:** October 27th, 2026

**DOI:** <https://doi.org/10.21203/rs.3.rs-10764908/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

**Additional Declarations: Yes** O.A.A. is a consultant to Precision-Health.ai and Ledidi, and has received speaker's honoraria from Eli Lilly and Company, Bristol Myers Squibb, H. Lundbeck A/S, Otsuka Pharmaceutical Co, Sunovion Pharmaceuticals Inc., and Janssen Pharmaceuticals.

---

# Abstract

Co-occurring internalizing and externalizing behavioral problems in childhood are associated with greater symptom persistence and increased psychiatric burden than either in isolation, yet their molecular correlates in early development remain poorly characterized. We analyzed behavioral profiles at ages three and five years in the Norwegian Mother, Father and Child Cohort Study (MoBa), using a multi-layer and multi-cohorts design integrating genome-wide association analysis, polygenic score analysis, epigenome-wide association analysis of cord blood DNA methylation, and co-methylation network analysis. Behavioral profiles were based on previously identified latent classes representing internalizing, aggression, and co-occurring problems at each age. Across six genome-wide association analyses, covering these three behavioral profiles at ages three and five, we identified one genome-wide significant locus for co-occurring problems at age five, led by rs73161798 at chromosome 13q12 ( $P = 3.26 \times 10^{-8}$ ). SNP-based heritability estimates were modest ( $h^2_{\text{SNP}} \approx 0.03\text{--}0.07$ ) but became more consistently detectable from age three to five. Polygenic score analyses provided the most consistent molecular signal, with broader associations at age five across psychiatric, cognitive, and neurodevelopmental domains. In contrast, single-site cord-blood DNA methylation associations were sparse, identifying one age-three cytosine-phosphate-guanine site (CpG) associated with co-occurring problems, annotated to *TNKS* (cg10976481;  $P = 9.03 \times 10^{-8}$ ), and three age-five CpGs, including one in the *TENM4* locus (cg18224993;  $P = 6.38 \times 10^{-8}$ ). These methylation associations were then examined in two independent cohorts, ALSPAC and GenR. Co-methylation network analysis identified age-five modules associated with co-occurring profiles, with nominal enrichment for pathways related to neurodevelopment, synaptic organization, neuronal signaling, and plasticity. These findings thus suggest that molecular correlates become more consistently detectable as behavioral phenotypes become more stable across early childhood from age three to five.

## Introduction

Childhood behavioral problems are common, developmentally heterogeneous, and clinically consequential<sup>1–3</sup>. Internalizing difficulties, such as anxiety, withdrawal and emotional symptoms, and externalizing behaviors, such as oppositionality and physical or relational aggression, are often studied as distinct domains<sup>4</sup>. In early childhood, however, these problems frequently co-occur within the same child<sup>5,6</sup>. This co-occurrence is clinically important because children exhibiting both internalizing and externalizing symptoms tend to experience more persistent difficulties and greater later psychiatric burden than children with single-domain symptom profiles<sup>7–11</sup>. Despite this, molecular studies of early behavioral problems have largely focused on individual symptom domains, leaving the genetic and epigenetic correlates of co-occurring behavioral profiles poorly understood.

A central unresolved question is whether co-occurring behavioral problems primarily reflect overlapping liabilities already implicated in internalizing and externalizing traits, or whether they also mark a more developmentally distinct profile<sup>12,13</sup>. Twin and genomic studies indicate that childhood psychopathology

is shaped by both shared and domain-specific genetic influences, and psychiatric genetic studies increasingly support broad pleiotropic liability across emotional, behavioral, cognitive, and neurodevelopmental traits<sup>14–20</sup>. At the same time, person-centered developmental studies show that co-occurring profiles can be identified early in life and are associated with elevated risk later in development<sup>21</sup>. These observations support a framework in which co-occurring behavioral problems are not viewed simply as the additive combination of internalizing and externalizing behavior, but rather as a potentially informative early-life phenotype that may index broader developmental vulnerability.

The timing of assessment is also critical<sup>10,22</sup>. Behavioral problems measured at in early childhood may partly reflect transient variation in early self-regulation, language development, temperament, and parent-reported adjustment<sup>23</sup>. By mid-childhood, behavioral profiles may be more stable and more closely aligned with emerging cognitive, emotional, and social regulation. Molecular correlates in early childhood may be subtle, yet they can provide early insight into emerging biological pathways that become more structured as behavioral phenotypes stabilize. This developmental perspective is particularly relevant for co-occurring problems, which may index broader liability more clearly once symptom patterns become more stable and differentiated. Common genetic variation provides one route to characterizing this liability. Genome-wide association analysis can identify individual loci associated with early behavioral profiles<sup>20,24</sup>, while single nucleotide polymorphism (SNP)-based heritability and polygenic score (PGS) analyses can test whether childhood behavioral profiles reflect broader genetic liabilities for psychiatric, cognitive, and neurodevelopmental traits. We therefore selected PGSs across psychiatric, cognitive, educational, and neurodevelopmental domains to examine whether early behavioral profiles show genetic continuity with traits and disorders that commonly emerge or become clinically salient later in development. However, individual common-variant effects for early childhood behavior are expected to be small, and genome-wide significant findings may be difficult to detect even in large cohorts. Polygenic approaches may therefore provide a more sensitive framework for evaluating whether early behavioral profiles are developmentally linked to later psychiatric and cognitive vulnerability.

DNA methylation offers a complementary molecular perspective. Cord-blood methylation is measured before the emergence of later childhood behavioral profiles and may capture prenatal biological variation related to genetic regulation, fetal development, maternal exposures, cell composition, and other early-life processes. However, methylation associations with behavioral phenotypes are often weak, context-dependent, and difficult to interpret at the level of single CpG sites. Network-level approaches, which model correlated patterns of CpG sites as interconnected modules rather than isolated loci, may therefore be useful for identifying coordinated methylation structure even when individual CpG associations are sparse. Such approaches can provide biological context, while still requiring cautious interpretation because cord blood is not brain tissue and methylation signatures do not by themselves establish causal mechanisms.

Here, we used the Norwegian Mother, Father and Child Cohort Study (MoBa)<sup>25,26</sup>, to characterize molecular correlates of early childhood internalizing, externalizing (indexed by aggression), and co-

occurring behavioral profiles at ages three and five years (Fig. 1). Our findings were then examined in two independent cohorts, the ALSPAC and Generation R described below.

Behavioral profiles were based on Tsotsi et al. (2023), who identified latent profiles reflecting primarily internalizing difficulties, primarily aggressive behavior, co-occurring internalizing difficulties and aggression, and low-symptom/normative development. In the present study, we refer to the primarily aggressive profile as the externalizing profile, consistent with this previously derived latent profile structure. We integrated genome-wide association analyses, SNP-based heritability estimation, polygenic score analyses, epigenome-wide association analyses of cord-blood DNA methylation, and weighted co-methylation network analysis. Rather than assuming that each molecular layer would yield equally strong evidence, we asked whether a coherent developmental pattern would emerge across genetic and epigenetic analyses. We hypothesized that molecular signals would be modest in early childhood and stronger by age five than by age three, with polygenic analyses providing the clearest evidence and cord-blood methylation contributing complementary, hypothesis-generating biological context.

## **Methods and Materials**

### **Sample and data processing**

#### **The Mother, Father, and Child Cohort study (MoBa)**

Participants were recruited from all over Norway from 1999 to 2008. The women consented to participation in 41% of pregnancies. The cohort includes approximately 113,632 children, 94,834 mothers and 75,229 fathers<sup>25</sup>. The present study focuses on child data, with umbilical cord blood samples collected at birth for most participants. This study is based on version 12 of the quality-assured MoBa dataset, released for research in 2021. The establishment of MoBa and initial data collection were approved by the Norwegian Data Protection Agency and the Regional Committees for Medical and Health Research Ethics (REK:185800). The MoBa cohort is currently regulated by the Norwegian Health Registry Act. This study received ethical approval from REK (2009/1899-7; 2013/2061) and the Norwegian Data Inspectorate, with informed consent obtained from all participants prior to participation.

#### **The Generation R Study**

The Generation R Study is a population-based cohort on health and development from fetal life onwards<sup>27,28</sup>. Pregnant women residing in the study area of Rotterdam, the Netherlands, with an expected delivery date between April 2002 and January 2006 were invited to participate. The study has been approved by the Medical Ethical Committee of Erasmus MC, University Medical Center Rotterdam. Written informed consent was obtained from parents of all children. In total, 9,778 mothers were enrolled in the study and gave birth to 9,749 live-born children. In two subsamples of children of parents born in

the Netherlands, the cord blood of 1,393 children was epigenotyped with the Illumina Infinium HumanMethylation450 BeadChip (Illumina Inc. Sand Diego, CA) and the cord blood of another 1,115 children was epigenotyped with the Illumina Infinium MethylationEPIC v1.0 BeadChip. Of these, 1,145 and 897 children had data available on the Child Behavior Checklist <sup>29</sup> at age 6 years (CBCL/6–18), respectively. Of these, we analyzed the data of the 1,076 and 843 children had complete data on all covariates, respectively. There were no siblings present within or between these datasets.

## The ALSPAC study

Pregnant women resident in Avon, UK with expected dates of delivery between 1st April 1991 and 31st December 1992 were invited to take part in the study. The initial number of pregnancies enrolled was 14,541. Of the initial pregnancies, there were 13,988 children who were alive at 1 year of age (Boyd et al., 2013; Fraser et al., 2013). The total sample size for analyses using any data collected after the age of seven is therefore 15,447 pregnancies. Of these 14,901 children were alive at 1 year of age. The study website contains details of all available data through a fully searchable data dictionary and variable search tool (<http://www.bristol.ac.uk/alspac/researchers/our-data/>). Our study population is drawn from the Accessible Resource for Integrated Epigenomics Studies (ARIES) project, a sub-study drawn from the ALSPAC mother-child cohort on a subset of 1,018 mother-child pairs, which has DNA methylation available. For inclusion, participants were required to have DNA methylation data at birth and Strengths and Difficulties Questionnaire (SDQ) data at age 4, resulting in a final sample size of 706 individuals. Ethical approval was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. Informed consent for the use of all data collected was obtained from participants following the recommendations of the ALSPAC Ethics and Law Committee at the time. Participants can contact the study team at any time to retrospectively withdraw consent for their data to be used. Study participation is voluntary and during all data collection sweeps, information was provided on the intended use of data.

## Measures

### MoBa

Child behavior was assessed using items from the Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2000), as described in Tsotsi et al. (2023). Mothers rated their children's behaviors on a three-point scale ranging from 0 (not true) to 2 (very true). In Tsotsi et al. (2023), exploratory and confirmatory factor analyses were first used to identify the internalizing difficulties and aggression indicators at ages three and five years. Internalizing difficulties were assessed with nine items at age three and ten items at age five, with raw scores ranging from 0 to 18 and 0 to 20, respectively. Aggression was assessed with four items at both ages, with raw scores ranging from 0 to 8.

These factor-derived internalizing and aggression scores were then entered into latent profile analyses to identify data-driven behavioral profiles at each age. The analyses identified four profiles: low-symptom/normative, primarily internalizing, primarily aggressive, and co-occurring internalizing difficulties and aggression. In the present study, we refer to the primarily aggressive profile as the externalizing profile, reflecting the externalizing-related manifestation captured by this previously derived profile structure.

Latent profile analysis was performed separately for ages three and five to identify behavioral subgroups. Based on model performance, four groups were identified at each age (**Suppl Figs. S1 and S2**): a group with low or normal levels of behavioral problems (normative group), a group with primarily internalizing difficulties (internalizing), a group with primarily externalizing behaviors (aggression), and a group with co-occurring internalizing and externalizing behaviors. The distribution of children across these groups at each age is shown in **Suppl Fig. S3**. Similar Latent profiles were developed in Gen R and ALSPAC.

## The Generation R Study

The Child Behavior Checklist (CBCL/6–18) was reported by the main caregiver at age 6<sup>29</sup>. Items were scored on a 3-point Likert scale (0 = not true, 1 = somewhat or sometimes true, and 2 = very true or often true). Weighted sumscores were computed for internalizing behavior (31 items comprising anxious/depressed, withdrawn/depressed and somatic complaints) and aggressive behavior (19 items), allowing for up to 25% missing.

## ALSPAC

The parent-reported SDQ was administered to children at a median age of four. The SDQ consists of five domains: emotional problems, peer problems, conduct problems, hyperactivity/inattention, and prosocial behavior. Each subscale consists of five items scored on a 3-point Likert-type scale (not true, sometimes true, and certainly true) with a response option for can't say/not applicable. Items were summed to generate an overall score for each domain that ranges from 0 to 10. In line with SDQ recommendations, we derived an 'Internalizing problems' variable by summing emotional problems and peer problems for a composite score. Externalising problems was indexed by conduct problems. A latent class analysis with a three-profile solution was carried out with these variables and the posterior probabilities of belonging to class 3 were used as the co-occurring phenotype used in EWAS analysis.

## Genotypes and Imputation

### MoBa

The MoBa genetics data is managed by a centralized core team, with strict quality control procedures applied before data sharing across subprojects. The present study is based on version v1 of the dataset, which was generated using a carefully designed pipeline that accounts for the unique family structure in the MoBa study while performing imputation<sup>30</sup>. This pipeline imputed and released 6,981,748 hard-coded autosomal SNPs for 207,569 unique individuals of European ancestry, along with sex chromosome genotypes for a subset of individuals. The present study did not use sex chromosome data or parental genotypes. Among the children with available behavioral data, up to 40,149 had quality-controlled genotype data and the exact number varies by behavioral and age group (**Suppl Fig. S3**).

Genotype-based principal components (PCs) were computed following standard procedures. Relatively rare SNPs, defined as those with a minor allele frequency (MAF) below 0.05, and SNPs failing Hardy-Weinberg equilibrium tests ( $P < 10^{-6}$ ) were removed. The remaining SNPs were thinned using a sliding window approach to remove highly correlated neighboring SNPs, as implemented in PLINK 2<sup>31</sup> with the parameters *-indep-pairwise* 100 50 0.1. The resulting thinned genotype data were then used as input for principal component analysis (PCA) using the PLINK *-pca* command. The top ten PCs were included in downstream analyses to account for subtle population stratification.

## DNA Methylation

### MoBa

DNA methylation profiles for MoBa participants were processed and quality-checked by a core team across ten data waves, following the pipeline documented at <https://github.com/folkehelseinstituttet/mobagen/wiki/Methylation#met001>. Briefly, preprocessing was conducted separately for each wave using the minfi R package<sup>32</sup>, following the 12-step protocol outlined in the pipeline.

The present study includes only child data and excludes waves designed for specific drug effects (waves 006 and 007, focused on Paracetamol) or those that included only ADHD patients. Waves 001 to 003 were based on the Illumina Infinium 450K BeadChip (Illumina, CA), while the remaining waves used the Illumina Infinium EPIC BeadChip (Illumina, CA). The quality-checked data were merged into a single dataset, retaining only CpG sites present on both array types. To mitigate batch effects, both wave indicators and chip type were included as covariates in all downstream analyses ( $n = 407,705$ ). In addition, CpG sites that showed low variation, std deviation  $< 0.01$ , were excluded and after these quality assurance steps about 196,794 CpG sites were analyzed for 6,074 individuals. As with genetic data, the number of unique individuals varies across behavioral groups (**Suppl Fig. S3**). EWAS results were examined for replication purposes in the Generation R and ALPSAC cohorts.

## The Generation R Study

DNA was extracted from cord blood samples taken at birth. Five-hundred nanograms of DNA per sample underwent bisulfite conversion using the EZ-96 DNA Methylation kit (Shallow) (Zymo Research

Corporation, Irvine, USA) and further processed with the Illumina Infinium HumanMethylation450 BeadChip (Illumina Inc., San Diego, USA) or Illumina Infinium MethylationEPIC v1.0 BeadChip. Preparation and normalization of the DNA methylation data was performed following the CPACOR workflow<sup>33</sup> using the software package R<sup>34</sup>. Arrays with observed technical problems such as failed bisulfite conversion, hybridization or extension as well as arrays with a sex mismatch were removed. Arrays with a call rate > 95% per sample were carried forward into normalization. The data were quantile normalized and represented as M-values.

## ALSPAC

DNA extracted from cord blood was bisulfite converted. Samples were processed with the Illumina Infinium HumanMethylation450 BeadChip (Illumina Inc., San Diego, CA). Quality control and EWAS were performed with the meffil package<sup>35</sup> on R Version 4.5.2. Samples with mismatched genotypes or sex, incorrect relatedness, low concordance with samples collected at other time points, extreme dye bias, or poor probe detection were excluded prior to normalization. Functional normalization was performed with 10 control probe principal components while including slide as a random effect. Probes were removed if they showed a detection  $p$ -value > 0.01 or a bead count below 3 in over 10% of samples. A total of 485,512 autosomal CpG sites passed these quality control steps.

## Statistical Analysis

### Genome-Wide Association Studies (GWAS)

Given the sample size constraints, we employed a case-control design instead of analyzing raw behavioral scores as continuous traits. Case status was defined from the previously established latent profile structure, which enabled data-driven identification of low-symptom/normative, primarily internalizing, primarily aggressive, and co-occurring profiles. These profiles have previously been shown to differentiate depressive, anxiety, conduct, and oppositional defiant symptoms at age eight, supporting their validity as meaningful early childhood risk profiles. We performed GWAS for three behavioral contrasts: internalizing versus low symptom group, aggression versus low, and co-occurring versus normal. Approximately 5% of the included children are siblings. To account for family structure, we used generalized linear mixed-effects models, where family identifiers were included as random intercept effects. Fixed effects included the SNP, the top ten PCs, sex, and genotyping batch indicators. The genetic relationship matrix was computed using the same thinned genotype data as for the PC analysis. GWAS was performed using GLMM models implemented in GCTA v1.94<sup>36</sup> with the command – *fastGWA-mlm-binary*. Genome-wide significance was defined as  $P < 5 \times 10^{-8}$ .

### SNP Heritability

The proportion of phenotypic variance attributable to genome-wide SNPs was estimated using GCTA's restricted maximum likelihood models. The significance of SNP-heritability estimates was assessed using the likelihood ratio test implemented in GCTA with the –*reml* command. To correct for multiple

testing, we applied the Benjamini-Hochberg false discovery rate (FDR), considering  $FDR \leq 0.05$  as statistically significant.

## Polygenic Risk Score (PGS) Analysis

To investigate genetic correlations between childhood behavioral problems and major neuropsychiatric conditions, cognitive traits, we conducted a PGS analysis. Summary statistics were obtained from GWAS on schizophrenia (SCZ) <sup>16</sup>, bipolar disorder (BIP) <sup>18</sup>, depression (DEP) <sup>37</sup>, autism spectrum disorder (ASD)<sup>38</sup>, attention deficit hyperactivity disorder (ADHD) <sup>39</sup>, generalized anxiety disorder (ANX) <sup>40</sup>, general intelligence (COG) <sup>41</sup>, educational attainment (EDU) <sup>42</sup>, neuroticism (NEU) <sup>43</sup>, and obsessive-compulsive disorder (OCD) <sup>44</sup>. As a negative control, we used hair color (HC) <sup>45</sup>. Summary statistics based on European samples were aligned to high-quality SNPs from the HapMap3 dataset before analysis. To account for winner's curse bias in SNP effect sizes, we applied a Bayesian mixture model, PRS-CS<sup>46</sup>. The posterior-corrected SNP effects were used as allele weights in computing PGS via the PLINK –score function.

We used multiple logistic regression models to assess the effect of PGS on the probability of exhibiting abnormal behavior compared to the normal group. The same covariates used in the GWAS analysis were included in the PGS models. To correct for multiple testing, we applied FDR procedure and results were considered statistically significant at  $FDR \leq 0.05$ .

## Epigenome-Wide Association Studies (EWAS)

We performed EWAS using the MOMENT model of the OSCA package in MoBa <sup>47</sup>, which, despite its conservative nature, effectively reduces batch effects by explicitly modeling neighboring methylation sites. Because DNA methylation levels can be either causal or consequential, this study tested whether methylation levels associate with the probability of being in an abnormal behavioral group compared to the normal group. In total, we analyzed 196,794 cytosine-phosphate-guanine (CpG) sites. To further minimize the risk of confounding, we included an extensive set of covariates, including gestational age at birth, maternal smoking during pregnancy (current, former, never), delivery mode (caesarean or vaginal), DNA methylation profiling wave and BeadChip type indicators, estimated immune cell proportions for six immune cell types, CD8 T cell, CD4 T cell, NK cell, monocyte, B cell and granulocyte, as well as DNA methylation based top ten principal components as covariates. We applied a p-value threshold of  $\leq 1 \times 10^{-7}$  to identify differentially methylated positions (DMPs).

To replicate significant DMPs identified in MoBa, logistic regression models were applied to the Generation R and ALSPAC DNA methylation data. The co-occurring profilevariable was used as dependent variable, and each of the significant DMPs was used as predictor, along with a set of covariates similar to that used in the MoBa analysis. These include the estimated proportions of the six immune cells, sex of child determined at birth, gestational age at birth, birth mode, maternal smoking during pregnancy. In addition, in the ALSPAC analysis, Surrogate variable analysis (SVA) resulted in 32

surrogate variables and was used to account for unmeasured sources of technical and biological variation.

## Variance in Behavioral Traits Explained by DNA Methylation

We estimated the proportion of variance in behavioral traits attributable to genome-wide DNA methylation variation using `–reml` function of the OSCA software. The same set of covariates used in the EWAS analysis was included in these models. Multiple testing correction was performed using the Benjamini-Hochberg FDR procedure, with  $FDR \leq 0.05$  considered statistically significant.

## Weighted Correlation Network Analysis (WGCNA)

We performed weighted correlation network analysis (WGCNA) <sup>48</sup> to explore the coordinated influence of DNA methylation on childhood behavioral problems. To reduce statistical noise and computational cost, we further restricted this analysis to CpGs that have a minimal standard deviation  $> 0.05$ , and medium absolute deviation  $>$  the 75% quantile. This reduced the number of CpGs from 196,794 down to 37,499.

We used the *pickSoftThreshold* function in the WGCNA package to determine the optimal power parameter for the scale-free network assumption, scanning values from 1 to 12. We decided the optimal power value to 2, which has an *SFT.R.sq* value of 0.937 and a *mean.k* 3.16. Based on this selected power, the *blockwiseModules* function was used to derive DNA methylation modules (MEs), with the following settings: *correlation type=bicor*, *minModuleSize* = 300, *TOMType='unsigned'*, *deepSplit* = 1, *maxBlockSize* = 3000, *mergeCutHeight* = 0.5. Other parameters were set to the default values. To summarize the functional relevance of these modules, eigengenes (the first principal component of the co-methylation networks) were computed using *moduleEigengenes*.

The association between eigengenes and behavioral groups was assessed using logistic regression, adjusting for the same covariates used in EWAS to control for confounding.

## Module Pathway Enrichment Analysis

We performed the pathway enrichment analysis for CpGs included in the DNA methylation modules, ME1 and ME3, using the standard package of in the field, *missMethyl* <sup>49</sup>. Pathways in both the KEGG <sup>50</sup> and GO BP databases (geneontology.org) were tested for enrichment. Since many gene sets within each database share genes and are therefore correlated, we applied the FDR procedure to correct for multiple testing, considering  $FDR \leq 0.05$  as the significance threshold.

## DNA Methylation Quantitative Trait Locus (mQTL) Analysis

For each identified DMP, we conducted cis-mQTL analysis using GCTA with the `–fastGWA-mlm` command. SNPs located within a 1 Mb radius of a DMP were analyzed. The same covariates used in the

GWAS analysis were included in the models to control for confounding effects. To correct for multiple testing, we applied the Bonferroni correction.

## Results

### Study overview

Using the MoBa cohort we examined internalizing, aggression, and co-occurring behavioral profiles at ages three and five years. Behavioral profiles were defined by latent profile analysis of CBCL<sup>29</sup> measures at each age, yielding typical, internalizing, aggression, and co-occurring groups. The distribution of children across these groups at each age is shown in Supplementary **Figs. S1** and **S2**. We then investigated genetic and epigenetic correlates of the three non-normative profiles relative to the normative group at each age (Fig. 1).

### Genetic contribution to Child Behavior Problems

Genome-wide SNP-based heritability ( $h^2_{\text{SNP}}$ ) estimates were modest overall. At age three, internalizing and aggression profiles showed significant estimates ( $h^2_{\text{SNP}} = 0.046, 0.061$ ; standard error [se] = 0.013, 0.013;  $FDR = 5.26 \times 10^{-4}, 6.06 \times 10^{-6}$ ; respectively), whereas co-occurring problem did not survive multiple-testing correction. By age five, all three behavioral phenotypes showed significant estimates  $h^2_{\text{SNP}}$ , with similar effect magnitudes: internalizing  $h^2_{\text{SNP}} = 0.033$  (se = 0.013,  $FDR = 1.151 \times 10^{-3}$ ), aggression  $h^2_{\text{SNP}} = 0.037$  (se = 0.014,  $FDR = 3.791 \times 10^{-3}$ ), and co-occurring  $h^2_{\text{SNP}} = 0.035$  (se = 0.015,  $FDR = 9.499 \times 10^{-3}$ ). These findings indicate that common variant contributions are detectable but modest and become more consistently evident at age five.

### Genetic Variants Associated with Child Behavioral Problems

Across the six GWAS performed, only co-occurring problems at age 5 yielded a genome-wide significant locus (**Suppl. Figs. S4-S7**). The lead variant, rs73161798, mapped to a gene-dense region on chromosome 13q12, with effect allele A showing  $\log(\text{OR}) = 1.08$  (se = 0.19,  $P = 3.26 \times 10^{-8}$ ) (Fig. 2A). Several genes with potential relevance to neurodevelopmental or behavioral biology were located in this locus, for example, the general transcription factor IIIA gene (*GTF3A*), the ubiquitin-specific peptidase 12 (*USP12*), and the Ras-like protein family member 11A (*RASL11A*). No comparable signal was observed for the other behavioral profiles or age groups. Genomic inflation was minimal ( $\lambda = 1.002-1.012$ ; **Suppl. Figs S5** and **S7**), supporting the robustness of the association.

### Polygenic score associations

Polygenic score analyses showed broader and more coherent associations at age five than at age three (Fig. 2B; see **Suppl Figs. S9, S10, Tables S1** and **S2** for details). Among psychiatric traits, neuroticism

PGS was associated with nearly all behavioral profiles across ages (FDR adjusted  $P < 0.05$ ), while educational attainment PGS generally showed inverse associations with behavioral risk. Internalizing problems at age five displayed the broadest pattern, being associated with all eight psychiatric PGSs, compared with only three at age three. One notable age-dependent pattern was observed for ADHD PGS: it was positively associated with aggression at age three ( $OR = 1.09$ , 95%CI: 1.05–1.134,  $FDR = 1.14 \times 10^{-4}$ ) but with internalizing at age five ( $OR = 1.21$ , 95%CI: 1.157–1.266,  $FDR = 2.22 \times 10^{-15}$ ), and negatively associated with aggression at age five ( $OR = 0.92$ , 95%CI: 0.881–0.965,  $FDR = 2.15 \times 10^{-3}$ ). Overall, these findings suggest that polygenic liability is developmentally structured and more detectable at age five.

## Longitudinal pattern across ages

Among children with behavioral data available at both ages, summary analyses showed that stable internalizing profiles carried the broadest polygenic burden across cognitive, neurodevelopmental, and psychiatric score domains (Fig. 2C). Stable aggression and co-occurring profiles also showed polygenic signal, but with less consistent cross-domain loading. These results support the view that persistence across early childhood may index a more structured genetic liability.

## DNA Methylation and childhood Behavioral Problems

We estimated the proportion of behavioral variation explained by epigenome-wide DNA methylation in umbilical cord blood. Only co-occurring problems at age five showed significant DNA methylation-associated variance (variance explained = 0.297,  $se = 0.071$ ). These findings suggest that emergent co-occurring behavioral problems may carry a measurable epigenetic signal, whereas other behavioral profiles appear to be more captured by genotype-based analysis. Full estimates for all traits are provided in **Supplementary Tables S3 and S4**.

## Single-site methylation associations are limited but support a developmental contrast

After quality control of the Illumina 450 and EPIC arrays and removal of sites that failed quality control or showed little variation, we performed genome-wide differential methylation analysis (EWAS) on 196,794 sites (**Methods**) in MoBa. At age three, cg10976481, annotated to the first exon of the Tankyrase gene (*TNKS*), was associated with the co-occurring problems ( $\beta = -0.024$ ,  $se = 0.004$ ,  $P = 9.03 \times 10^{-8}$ ) (Fig. 3A). At age five, we identified three independent differentially methylated sites (Fig. 3B-D): cg18224993 in the enhancer region of the Teneurin transmembrane protein 4 (*TENM4*, previously named ODZ4;  $\beta = -0.03$ ,  $se = 0.003$ ,  $P = 6.38 \times 10^{-8}$ ), cg03585974 near the transmembrane protein 239 (*TMEM239*;  $\beta = -0.027$ ,  $se = 0.005$ ,  $P = 7.31 \times 10^{-9}$ ), and the intergenic site cg03106395 ( $\beta = -0.028$ ,  $se = 0.005$ ,  $P = 2.21 \times 10^{-8}$ ). We did not identify significant sites for internalizing or aggression symptom profiles. Inflation statistics did not indicate substantial bias due to cryptic relatedness nor unmodelled

confounding ( $\lambda$ : 0.82–0.98; **Suppl Figs. S12–16**), and the volcano plots showed no obvious abnormalities.

We next examined these four differentially methylated sites in two external cohorts, the Generation R and the ALSPAC datasets (**Methods**). In Generation R, cg18224993 in the *TENM4* locus showed consistent direction of effect and a nominal association with the co-occurring problems (z-score = -2.053,  $P = 0.04$ ) with the Illumina EPIC array data ( $n = 843$ ). In the Illumina 450K data ( $n = 1,145$ ) this probe showed consistent effect direction but not significant ( $P > 0.05$ ). In ALSPAC, none of the four sites showed nominal evidence of associations (all  $P > 0.05$ ;  $n = 706$ ). Taken together, these external lookups provide partial support for the *TENM4* signal in Generation R, while indicating that the single-site methylation findings should be interpreted cautiously and require further evaluation in larger harmonized samples.

Overall, the single-site methylation findings were modest and focal rather than widespread (Fig. 3E). The age-five co-occurring profile showed the clearest methylation signal, including a CpG in the *TENM4* locus with partial external support in Generation R but not ALSPAC. As in the genetic analyses above, the contrast between ages three and five was notable: methylation signals were weaker and scattered at age three, but more structured and biologically interpretable at age five.

## Co-methylation network analysis identifies biologically coherent modules at age five

Because single-site methylation results were limited, we asked whether coordinated methylation structure could be detected at the network level. Weighted co-methylation network analysis identified five eigengene modules (ME1-5, Fig. 4A). Correlations among the eigengenes suggested three separable module groups, with ME1 and ME2 negatively correlated ( $r = -0.35$ ), ME4 and ME5 positively correlated ( $r = 0.43$ ), and ME3 itself as a unique group, indicating that these groups may capture distinct biological processes. Using logistic regression while controlling relevant covariates, we found that at age five, ME1 was associated with co-occurring problems ( $P = 0.045$ ) and ME3 with aggression ( $P = 3.73 \times 10^{-3}$ ). In contrast, age-three module associations were weaker and less specific (see **Suppl. Table S5** for full results).

Although pathway enrichment for the methylation modules did not survive multiple-testing, nominally enriched pathways were biologically coherent (Fig. 4C; **Suppl. Tables S6–S9**). For ME1, selected terms converged on neurodevelopmental and synaptic themes, including neuron development, dendrite development, neuron projection morphogenesis, synapse organization, axon guidance, glutamatergic synapse, and related signaling pathways. For ME3, nominal enrichment highlighted pathways related to neuronal plasticity and signaling, including collateral sprouting, presynaptic membrane regulation, GABAergic synapse, and lipid- or signaling-related processes.

These results do not provide definitive pathway-level evidence. Rather, they offer hypothesis-generating biological context that is consistent with the broader developmental pattern observed across the study

## Discussion

In this multi-layer molecular analysis of early childhood behavioral profiles in MoBa, we found that molecular correlates of internalizing, aggressive, and co-occurring problems were modest overall but become more coherent by age five. The strongest and most consistent evidence came from polygenic score analyses, which showed broader associations at age five than at age three across psychiatric, cognitive, and neurodevelopmental domains. GWAS identified one genome-wide significant locus for co-occurring problems at age five, whereas single-site cord-blood methylation associations were limited. However, at age five co-methylation network analysis identified modules associated with co-occurring and aggression profiles, with nominal enrichment for neurodevelopmental, synaptic, and neuronal-plasticity themes. Together, these findings suggest that molecular correlates of early behavioral vulnerability are detectable but modest and become more coherent from age three to five.

We identified a genome-wide significant association at chr13q12 for co-occurring problems at age five. The region contains several biologically plausible nearby genes, including *GTF3A*, *USP12*, and *RASL11A*, but the present data do not permit prioritization of a single causal gene. *GTF3A* has been implicated in cellular responses relevant to viral defense in the brain<sup>51</sup>. *USP12* has been linked to adult risk-taking behavior<sup>20</sup>, and *RASL11A* has been discussed in relation to neurodevelopmental processes<sup>52</sup>. These annotations provide biological plausibility, but the association cannot yet be assigned to a specific causal gene. The locus therefore requires replication in independent cohorts with comparable early-childhood behavioral profiling, followed by functional annotation to clarify the underlying gene-regulatory mechanism.

The polygenic score analyses provided the clearest evidence that early behavioral profiles index broader psychiatric and neurodevelopmental liability. Internalizing problems at age five showed the broadest pattern of association across psychiatric PGSs, consistent with prior evidence for shared genetic architecture across depression, anxiety, neuroticism, and related internalizing traits<sup>53</sup>. The age-dependent ADHD PGS pattern was particularly notable. Higher ADHD PGS was associated with aggression at age three, but by age five it was positively associated with internalizing problems and inversely associated with aggression. One possible interpretation is that genetic liability related to ADHD may manifest differently across early development: at younger ages it may be expressed primarily through impulsivity, poor inhibitory control, or dysregulated behavior, whereas at older ages it may increasingly be reflected in emotional dysregulation, anxiety-related behavior, or broader internalizing vulnerability. This interpretation is consistent with evidence that early externalizing, such as aggression, and internalizing symptoms are dynamically related across childhood rather than developmentally fixed<sup>54–56</sup>. More generally, these findings suggest that the behavioral expression of polygenic liability may shift with age, underscoring the importance of repeated developmental assessment<sup>57</sup>. Alternatively, changing associations may, at least partly, reflect developmental differences in symptom measurement or phenotypic composition rather than changes in the expression of genetic liability.

Cord-blood DNA methylation contributed a complementary but more exploratory layer of evidence. Because methylation was measured at birth, these signals preceded the behavioral assessments and may reflect prenatal biological variation. However, cord blood is not brain tissue, and these findings cannot establish a direct causal mechanism linking prenatal exposures to later behavior. The sparse single-site EWAS results are therefore best interpreted as focal molecular correlates rather than evidence for a broad methylation architecture of early behavioral problems. The age-five *TENM4*-associated CpG is notable because teneurin-family proteins are conserved cell-adhesion molecules with established roles in axon guidance, synapse formation, neuronal recognition, and circuit wiring<sup>58,59</sup>, while *TENM4/ODZ4* has also been implicated in psychiatric genetics<sup>60</sup>. Nevertheless, the lack of replication in ALSPAC and Gen R underscores the need for larger samples and tissue-aware functional annotation before strong mechanistic conclusions can be drawn. The co-methylation network results suggest that coordinated methylation patterns may provide biological context beyond single CpG associations. ME1, associated with age-five co-occurring problems, showed nominal enrichment for neurodevelopmental and synaptic pathways, including neuron development, dendrite development, axon guidance, glutamatergic synapse, and long-term potentiation. ME3, associated with age-five aggression, highlighted more mixed but still biologically plausible themes, including GABAergic synapse, presynaptic membrane regulation, cytoskeletal organization, cerebellar development, cholesterol metabolism, and xenobiotic response. Because these enrichments did not survive multiple-testing correction, they should not be interpreted as definitive pathway discoveries. Instead, they support a hypothesis-generating model in which age-five behavioral profiles are accompanied by coordinated cord-blood methylation variation related to neurodevelopmental and neuronal-signaling processes.

Several aspects strengthen this study. First, we used a large, population-based birth cohort with behavioral assessments at two early developmental stages, enabling us to compare molecular correlates at ages three and five. Second, the same broad behavioral domains were available across ages, allowing us to evaluate whether molecular signals became more structured across early childhood. Third, the integration of GWAS, SNP-based heritability, polygenic scores, EWAS, and co-methylation network analysis allowed us to compare multiple molecular layers rather than relying on a single analytic framework. Finally, the use of cord-blood methylation provided a temporally informative measure collected before the emergence of the behavioral profiles studied here.

Several limitations should also be noted. First, behavioral profiles were derived from categorical latent profile assignments, which may simplify the dimensional and probabilistic nature of early behavioral variation<sup>10</sup>. Although this approach allowed us to contrast co-occurring and domain-specific profiles against a normative group, uncertainty in profile assignment was not fully propagated into the molecular analyses<sup>61</sup>. Second, behavioral profiles also relied on maternal report and may therefore partly reflect reporter-specific variance. Third, despite the large cohort setting, the number of children in non-normative behavioral groups was more limited than the full MoBa sample, reducing power for GWAS, EWAS, and subgroup-specific analyses. Third, the chr13q12 GWAS locus and methylation findings require independent replication with large samples. Fourth, cord-blood methylation is an accessible

early-life tissue but does not directly index methylation patterns in the developing brain. Finally, pathway enrichment results were nominal and should be interpreted as hypothesis-generating rather than confirmatory.

In conclusion, this study supports a developmental, multi-layer view of early childhood behavioral vulnerability (Fig. 5). Co-occurring behavioral problems were not characterized by a single dominant molecular signature, but by modest signals distributed across genetic and epigenetic layers, most clearly detectable by age five. Polygenic score analyses provided the strongest evidence for shared psychiatric, cognitive, and neurodevelopmental liability, whereas cord-blood methylation contributed complementary but more exploratory biological context through sparse CpG findings and network-level structure. These results highlight the value of integrating behavioral profiling with genomic and epigenomic data in early childhood, while underscoring the need for replication, longer follow-up, and tissue-aware interpretation of methylation findings.

## Declarations

### Acknowledgements

We are grateful to all the participating families in Norway who take part in this on-going cohort study. For generating high-quality genomic data, we thank the Norwegian Institute of Public Health (NIPH), the HARVEST collaboration, the NORMENT Centre at the University of Oslo, the Center for Diabetes Research at the University of Bergen, deCODE Genetics, the Research Council of Norway, the South-Eastern and Western Norway Regional Health Authorities, the ERC AdG, Stiftelsen KG Jebsen, the Trond Mohn Foundation, and the Novo Nordisk Foundation. Study funding: The Norwegian Mother, Father and Child Cohort study is supported by the Norwegian Ministry of Health and the Norwegian Research Council/FUGE (grant number 151918/S10). This work was also supported by the Research Council of Norway grant number 288083 and 301004 and by the Research Council of Norway FRIPRO (no.302854). In addition, LJH is supported by funding from the Norwegian South-East Regional Health Authority (Helse South-East;2022083).

The Generation R Study is conducted by Erasmus MC, University Medical Center Rotterdam in close collaboration with the School of Law and Faculty of Social Sciences of Erasmus University Rotterdam, the Municipal Health Service Rotterdam area, Rotterdam, the Rotterdam Homecare Foundation, Rotterdam and the Stichting Trombosedienst & Artsenlaboratorium Rijnmond (STAR-MDC), Rotterdam. We gratefully acknowledge the contribution of children and parents, general practitioners, hospitals, midwives and pharmacies in Rotterdam. The generation and management of the Illumina 450K methylation array data (EWAS data) for the Generation R Study was executed by the Human Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, the Netherlands. We thank Mr. Michael Verbiest, Ms. Mila Jhamai, Ms. Sarah Higgins, Mr. Marijn Verkerk and Dr. Lisette Stolk for their help in creating the EWAS database. We thank Dr. A. Teumer for his work on the quality control and normalization scripts. The general design of The Generation R Study was made possible by

financial support from Erasmus MC, Rotterdam, Erasmus University Rotterdam, the Netherlands Organization for Health Research and Development (ZonMW), the Netherlands Organization for Scientific Research (NWO), the Ministry of Health, Welfare and Sport and the Ministry of Youth and Families. The EWAS data were funded by a grant from the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) Netherlands Consortium for Healthy Aging (NCHA; project nr. 050-060-810), by funds from the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, by a grant from the National Institute of Child and Human Development (R01HD068437), and by a Helmholtz International Fellow Award HIFA-0174 -RA-30/19. The work of CAMC is supported by the European Union's HorizonEurope Research and Innovation Programme (FAMILY, grant agreement No 101057529; HappyMums, grant agreement No 101057390) and the European Research Council (TEMPO; grant agreement No 101039672).

We are also extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes data collection staff, data and administrations staff, technical managers and the technical staff with the Bristol Bioresource Laboratory, based within the University of Bristol. The UK Medical Research Council and Wellcome (Grant ref: MR/Z505924/1) and the University of Bristol provide core support for ALSPAC. This publication is the work of the authors and YW will serve as guarantors for the contents of this paper. ZH is supported by the Economic and Social Research Council (ESRC). EW and VB are funded from UK Research and Innovation (UKRI) under the UK government's Horizon Europe / ERC Frontier Research Guarantee [BrainHealth, grant number EP/Y015037/1]

## **Conflict of Interest**

O.A.A. is a consultant to Precision-Health.ai and Ledidi, and has received speaker's honoraria from Eli Lilly and Company, Bristol Myers Squibb, H. Lundbeck A/S, Otsuka Pharmaceutical Co, Sunovion Pharmaceuticals Inc., and Janssen Pharmaceuticals.

## **Data availability**

Genome-wide summary statistics for child behavior problems are available at <http://figshare/XXXX>. GWAS summary statistics for psychiatric disorders (SCZ, BIP, ASD, ADHD, PTSD, DEP, ANX, and OCD) can be downloaded from <https://pgc.unc.edu/for-researchers/download-results/> and for IQ and neuroticism from [https://cncr.nl/research/summary\\_statistics/](https://cncr.nl/research/summary_statistics/). Data on educational attainment

can be found at <http://www.thessgac.org/>.

## **Code availability**

Genome-wide association analysis for child behavior problems was performed using GCTA v1.94 (<https://yanglab.westlake.edu.cn/software/gcta/#fastGWA>). Polygenic score analysis was conducted with PRS-CS (<https://github.com/getian107/PRSCs>) and PLINK 2.0 (<https://www.cog-genomics.org/plink/2.0/>). EWAS analysis was performed using the OSCA toolset

(<https://yanglab.westlake.edu.cn/software/osca>). Gene set analysis used the WGCNA R package (<https://cran.r-project.org/web/packages/WGCNA/index.html>).

## References

1. Campbell, S. B. in *Advances in Clinical Child Psychology* (eds Thomas H. Ollendick & Ronald J. Prinz) 1–26 (Springer US, 1997).
2. Vergunst, F. *et al.* Association of Childhood Externalizing, Internalizing, and Comorbid Symptoms With Long-term Economic and Social Outcomes. *JAMA Network Open* 6, e2249568–e2249568 (2023). <https://doi.org/10.1001/jamanetworkopen.2022.49568>
3. Arslan, İ. B., Lucassen, N., van Lier, P. A. C., de Haan, A. D. & Prinzie, P. Early childhood internalizing problems, externalizing problems and their co-occurrence and (mal)adaptive functioning in emerging adulthood: a 16-year follow-up study. *Social psychiatry and psychiatric epidemiology* 56, 193–206 (2021). <https://doi.org/10.1007/s00127-020-01959-w>
4. Achenbach, T. M. *Assessment and taxonomy of child and adolescent psychopathology*. (Sage Publications, Inc, 1985).
5. Angold, A., Costello, E. J. & Erkanli, A. Comorbidity. *The Journal of Child Psychology and Psychiatry and Allied Disciplines* 40, 57–87 (1999). <https://doi.org/10.1111/1469-7610.00424>
6. Bista, S. *et al.* Joint developmental trajectories of internalizing and externalizing problems from mid-childhood to late adolescence and childhood risk factors: Findings from a prospective pre-birth cohort. *Development and psychopathology* 37, 176–191 (2025). <https://doi.org/10.1017/S0954579423001505>
7. Newman, D. L., Moffitt, T. E., Caspi, A. & Silva, P. A. Comorbid mental disorders: Implications for treatment and sample selection. *Journal of Abnormal Psychology* 107, 305–311 (1998). <https://doi.org/10.1037/0021-843X.107.2.305>
8. Basten, M. M. G. J. *et al.* The Dysregulation Profile in Young Children: Empirically Defined Classes in the Generation R Study. *Journal of the American Academy of Child & Adolescent Psychiatry* 52, 841–850.e842 (2013). <https://doi.org/10.1016/j.jaac.2013.05.007>
9. Shi, Q., Ettekal, I., Deutz, M. H. F. & Woltering, S. Trajectories of pure and co-occurring internalizing and externalizing problems from early childhood to adolescence: Associations with early childhood individual and contextual antecedents. *Developmental Psychology* 56, 1906–1918 (2020). <https://doi.org/10.1037/dev0001095>
10. Tsotsi, S. *et al.* Co-occurrence of internalizing difficulties and aggression in early childhood and risk of mental health problems in middle childhood. *International Journal of Behavioral Development* 47, 384–396 (2023). <https://doi.org/10.1177/01650254231187902>
11. Blok, E. *et al.* Stability and Change of Psychopathology Symptoms Throughout Childhood and Adolescence. *Child Psychiatry & Human Development* 53, 1330–1339 (2022). <https://doi.org/10.1007/s10578-021-01212-8>

12. Caspi, A. et al. The p Factor: One General Psychopathology Factor in the Structure of Psychiatric Disorders? *Clinical Psychological Science* 2, 119–137 (2013).  
<https://doi.org/10.1177/2167702613497473>
13. Forbes, M. K., Tackett, J. L., Markon, K. E. & Krueger, R. F. Beyond comorbidity: Toward a dimensional and hierarchical approach to understanding psychopathology across the life span. *Development and psychopathology* 28, 971–986 (2016). <https://doi.org/10.1017/S0954579416000651>
14. Grotzinger, A. D. et al. Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nature Human Behaviour* 3, 513–525 (2019).  
<https://doi.org/10.1038/s41562-019-0566-x>
15. Mullins, N. et al. Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nature Genetics* 53, 817–829 (2021).  
<https://doi.org/10.1038/s41588-021-00857-4>
16. Trubetskoy, V. et al. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* 604, 502–508 (2022). <https://doi.org/10.1038/s41586-022-04434-5>
17. Andreassen, O. A., Hindley, G. F. L., Frei, O. & Smeland, O. B. New insights from the last decade of research in psychiatric genetics: discoveries, challenges and clinical implications. *World psychiatry: official journal of the World Psychiatric Association (WPA)* 22, 4–24 (2023).  
<https://doi.org/https://doi.org/10.1002/wps.21034>
18. O'Connell, K. S. et al. Genomics yields biological and phenotypic insights into bipolar disorder. *Nature* 639, 968–975 (2025). <https://doi.org/10.1038/s41586-024-08468-9>
19. Cheesman, R. et al. Childhood behaviour problems show the greatest gap between DNA-based and twin heritability. *Translational psychiatry* 7, 1284 (2017). <https://doi.org/10.1038/s41398-017-0046-x>
20. Baselmans, B. et al. The Genetic and Neural Substrates of Externalizing Behavior. *Biological Psychiatry Global Open Science* 2, 389–399 (2022).  
<https://doi.org/https://doi.org/10.1016/j.bpsgos.2021.09.007>
21. Isdahl-Troye, A., Villar, P., Domínguez-Álvarez, B., Romero, E. & Deater-Deckard, K. The Development of Co-Occurrent Anxiety and Externalizing Problems from Early Childhood: a Latent Transition Analysis Approach. *Research on Child and Adolescent Psychopathology* 50, 505–519 (2022).  
<https://doi.org/10.1007/s10802-021-00865-2>
22. Tsotsi, S., Coplan, R. J., Czajkowski, N. O. & Bekkhus, M. Co-occurrence of internalizing difficulties and aggression in preschoolers and mental health outcomes in adolescence. *International Journal of Behavioral Development*, 01650254261438685 (2026). <https://doi.org/10.1177/01650254261438685>
23. Montroy, J. J., Bowles, R. P., Skibbe, L. E., McClelland, M. M. & Morrison, F. J. The development of self-regulation across early childhood. *Dev Psychol* 52, 1744–1762 (2016).  
<https://doi.org/10.1037/dev0000159>
24. Jami, E. S. et al. Genome-wide Association Meta-analysis of Childhood and Adolescent Internalizing Symptoms. *Journal of the American Academy of Child & Adolescent Psychiatry* 61, 934–945 (2022).  
<https://doi.org/https://doi.org/10.1016/j.jaac.2021.11.035>

25. Magnus, P. et al. Cohort Profile Update: The Norwegian Mother and Child Cohort Study (MoBa). *International journal of epidemiology* 45, 382–388 (2016). <https://doi.org/10.1093/ije/dyw029>
26. Brandlistuen, R. E. et al. Cohort Profile Update: The Norwegian Mother, Father and Child Cohort (MoBa). *International journal of epidemiology* 54, dyaf139 (2025). <https://doi.org/10.1093/ije/dyaf139>
27. Kooijman, M. N. et al. The Generation R Study: design and cohort update 2017. *European journal of epidemiology* 31, 1243–1264 (2016).
28. Kruithof, C. J. et al. The generation R study: Biobank update 2015. *European journal of epidemiology* 29, 911–927 (2014).
29. Achenbach, T. M. & Rescorla, L. A. *Manual for the ASEBA preschool forms and profiles*. Vol. 30 (Burlington, VT: University of Vermont, Research center for children, youth...
30. Corfield, E. C. et al. The Norwegian Mother, Father, and Child cohort study (MoBa) genotyping data resource: MoBaPsychGen pipeline v.1. *bioRxiv*, 2022.2006.2023.496289 (2024). <https://doi.org/10.1101/2022.06.23.496289>
31. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4, 7 (2015). <https://doi.org/10.1186/s13742-015-0047-8>
32. Aryee, M. J. et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* 30, 1363–1369 (2014). <https://doi.org/10.1093/bioinformatics/btu049>
33. Lehne, B. et al. A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. *Genome Biology* 16, 37 (2015).
34. R Core Team. R: A language and environment for statistical computing. (2013).
35. Min, J. L., Hemani, G., Davey Smith, G., Relton, C. & Suderman, M. Meffil: efficient normalization and analysis of very large DNA methylation datasets. *Bioinformatics* 34, 3983–3989 (2018). <https://doi.org/10.1093/bioinformatics/bty476>
36. Jiang, L., Zheng, Z., Fang, H. & Yang, J. A generalized linear mixed model association tool for biobank-scale data. *Nature Genetics* 53, 1616–1621 (2021). <https://doi.org/10.1038/s41588-021-00954-4>
37. Adams, M. J. et al. Trans-ancestry genome-wide study of depression identifies 697 associations implicating cell types and pharmacotherapies. *Cell* 188, 640–652.e649 (2025). <https://doi.org/10.1016/j.cell.2024.12.002>
38. Grove, J. et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat Genet* 51, 431–444 (2019). <https://doi.org/10.1038/s41588-019-0344-8>
39. Demontis, D. et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat Genet* 51, 63–75 (2019). <https://doi.org/10.1038/s41588-018-0269-7>

40. Strom, N. I. et al. Genome-wide association study of major anxiety disorders in 122,341 European-ancestry cases identifies 58 loci and highlights GABAergic signaling. *Nature Genetics* 58, 275–288 (2026). <https://doi.org/10.1038/s41588-025-02485-8>
41. Davies, G. et al. Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nat Commun* 9, 2098 (2018). <https://doi.org/10.1038/s41467-018-04362-x>
42. Lee, J. J. et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics* 50, 1112–1121 (2018). <https://doi.org/10.1038/s41588-018-0147-3>
43. Nagel, M. et al. Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. *Nature Genetics* 50, 920–927 (2018). <https://doi.org/10.1038/s41588-018-0151-7>
44. Strom, N. I. et al. Genome-wide analyses identify 30 loci associated with obsessive–compulsive disorder. *Nature Genetics* 57, 1389–1401 (2025). <https://doi.org/10.1038/s41588-025-02189-z>
45. Hysi, P. G. et al. Genome-wide association meta-analysis of individuals of European ancestry identifies new loci explaining a substantial fraction of hair color variation and heritability. *Nature Genetics* 50, 652–656 (2018). <https://doi.org/10.1038/s41588-018-0100-5>
46. Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C. A. & Smoller, J. W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nature Communications* 10, 1776 (2019). <https://doi.org/10.1038/s41467-019-09718-5>
47. Zhang, F. et al. OSCA: a tool for omic-data-based complex trait analysis. *Genome biology* 20, 107 (2019). <https://doi.org/10.1186/s13059-019-1718-z>
48. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9, 559 (2008). <https://doi.org/10.1186/1471-2105-9-559>
49. Phipson, B., Maksimovic, J. & Oshlack, A. missMethyl: an R package for analyzing data from Illumina’s HumanMethylation450 platform. *Bioinformatics* 32, 286–288 (2016). <https://doi.org/10.1093/bioinformatics/btv560>
50. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* 28, 27–30 (2000). <https://doi.org/10.1093/nar/28.1.27>
51. Naesens, L. et al. GTF3A mutations predispose to herpes simplex encephalitis by disrupting biogenesis of the host-derived RIG-I ligand RNA5SP141. *Science Immunology* 7, eabq4531 <https://doi.org/10.1126/sciimmunol.abq4531>
52. Rai, B. et al. Novel effects of Ras-MAPK pathogenic variants on the developing human brain and their link to gene expression and inhibition abilities. *Translational psychiatry* 13, 245 (2023). <https://doi.org/10.1038/s41398-023-02504-4>
53. Grotzinger, A. D. et al. Genetic architecture of 11 major psychiatric disorders at biobehavioral, functional genomic and molecular genetic levels of analysis. *Nature Genetics* 54, 548–559 (2022). <https://doi.org/10.1038/s41588-022-01057-4>

54. Jensen, C. M. & Steinhausen, H.-C. Comorbid mental disorders in children and adolescents with attention-deficit/hyperactivity disorder in a large nationwide study. *ADHD Attention Deficit and Hyperactivity Disorders* 7, 27–38 (2015). <https://doi.org/10.1007/s12402-014-0142-1>
55. Simonoff, E. et al. Psychiatric Disorders in Children With Autism Spectrum Disorders: Prevalence, Comorbidity, and Associated Factors in a Population-Derived Sample. *Journal of the American Academy of Child & Adolescent Psychiatry* 47, 921–929 (2008). <https://doi.org/https://doi.org/10.1097/CHI.0b013e318179964f>
56. Sterba, S. K., Prinstein, M. J. & Cox, M. J. Trajectories of internalizing problems across childhood: Heterogeneity, external validity, and gender differences. *Development and psychopathology* 19, 345–366 (2007). <https://doi.org/10.1017/S0954579407070174>
57. Askelund, A. D. et al. The Genetic Architecture of Differentiating Behavioral and Emotional Problems in Early Life. *Biological Psychiatry* 97, 1163–1174 (2025). <https://doi.org/https://doi.org/10.1016/j.biopsych.2024.12.021>
58. Li, J. et al. Structural Basis for Teneurin Function in Circuit-Wiring: A Toxin Motif at the Synapse. *Cell* 173, 735–748.e715 (2018). <https://doi.org/10.1016/j.cell.2018.03.036>
59. Gogou, C. et al. Alternative splicing controls teneurin-3 compact dimer formation for neuronal recognition. *Nature Communications* 15, 3648 (2024). <https://doi.org/10.1038/s41467-024-47763-x>
60. Sklar, P. et al. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nature genetics* 43, 977–983 (2011).
61. Nylund-Gibson, K., P., G. R. & and Masyn, K. E. Prediction from Latent Classes: A Demonstration of Different Approaches to Include Distal Outcomes in Mixture Models. *Structural Equation Modeling: A Multidisciplinary Journal* 26, 967–985 (2019). <https://doi.org/10.1080/10705511.2019.1590146>

## Figures

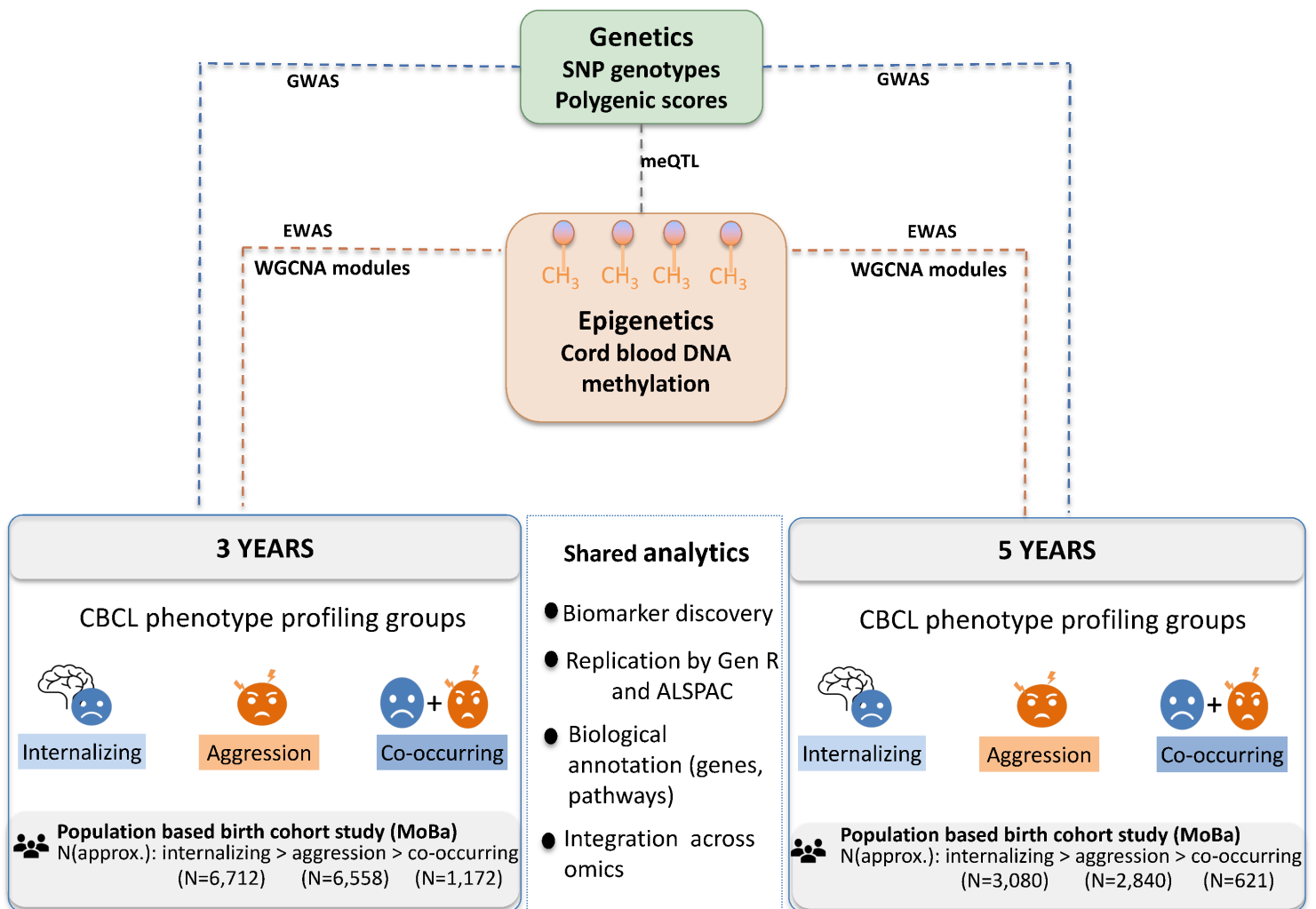


**Figure 1**

### **Study design and multi-layer molecular analysis of early childhood behavioral profiles.**

Internalizing, aggression, and co-occurring behavioral profiles were defined using Child Behavior Checklist (CBCL)-based phenotype profiling at ages three and five years in the Norwegian Mother, Father and Child Cohort Study (MoBa). At age three, the approximate sample sizes were 6,712 for internalizing, 6,558 for aggression, and 1,172 for co-occurring problems. At age five, the approximate sample sizes were 3,080 for internalizing, 2,840 for aggression, and 621 for co-occurring problems. Genetic analyses included SNP genotypes, genome-wide association analysis (GWAS), and polygenic score analyses. Epigenetic analyses used cord-blood DNA methylation and included epigenome-wide association analysis (EWAS), methylation quantitative trait locus (meQTL) information, and weighted co-methylation network analysis (WGCNA). Shared downstream analyses included biomarker discovery, biological annotation of genes and pathways, and integration across molecular layers.

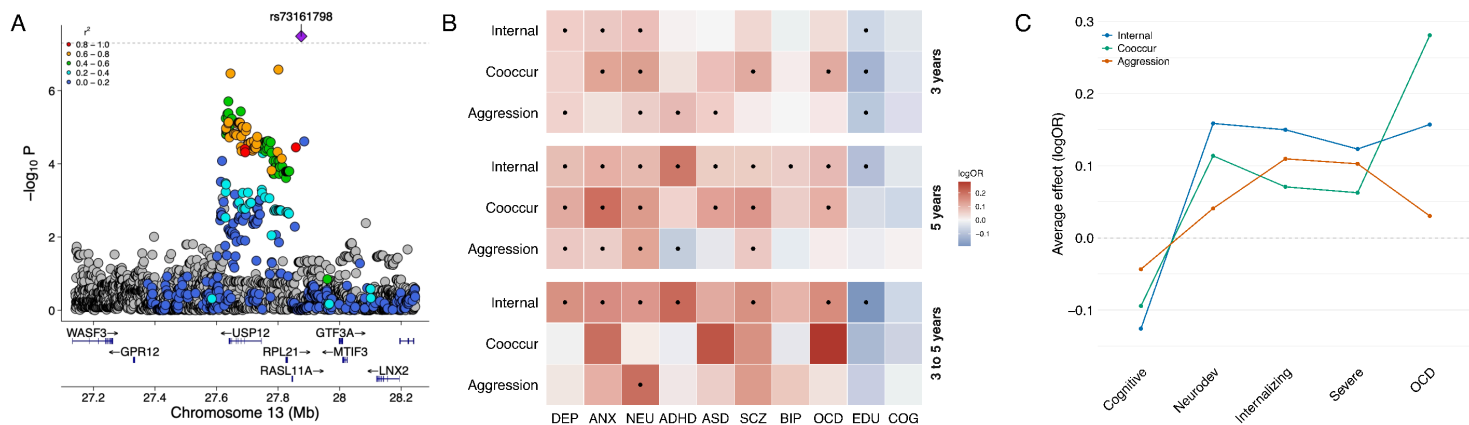


**Figure 2**

**Genetic findings for childhood behavioral profiles.**

**(A)** Regional association plot for the genome-wide significant locus associated with co-occurring problems at age five. The lead variant, rs73161798, maps to chromosome 13q12, with nearby genes including WASF3, GPR12, USP12, RPL21, RASL11A, GTF3A, MTIF3, and LNX2. Points are colored according to linkage disequilibrium with the lead variant. The dashed horizontal line indicates the genome-wide significance threshold. **(B)** Polygenic score associations across behavioral profiles and ages. Heatmap colors represent log odds ratios for associations between polygenic scores and internalizing, co-occurring, and aggression profiles at age three, age five, and persistent or transitional profiles from age three to age five. Black dots indicate associations passing multiple-testing correction. Polygenic scores include depression (DEP), anxiety (ANX), neuroticism (NEU), attention-deficit/hyperactivity disorder (ADHD), autism spectrum disorder (ASD), schizophrenia (SCZ), bipolar disorder (BIP), obsessive-compulsive disorder (OCD), educational attainment (EDU), and cognitive performance (COG). **(C)** Average polygenic score effects across broader score domains for longitudinal behavioral profiles. Values represent average log odds ratios across cognitive, neurodevelopmental, internalizing, severe psychiatric, and obsessive-compulsive disorder-related polygenic score domains.

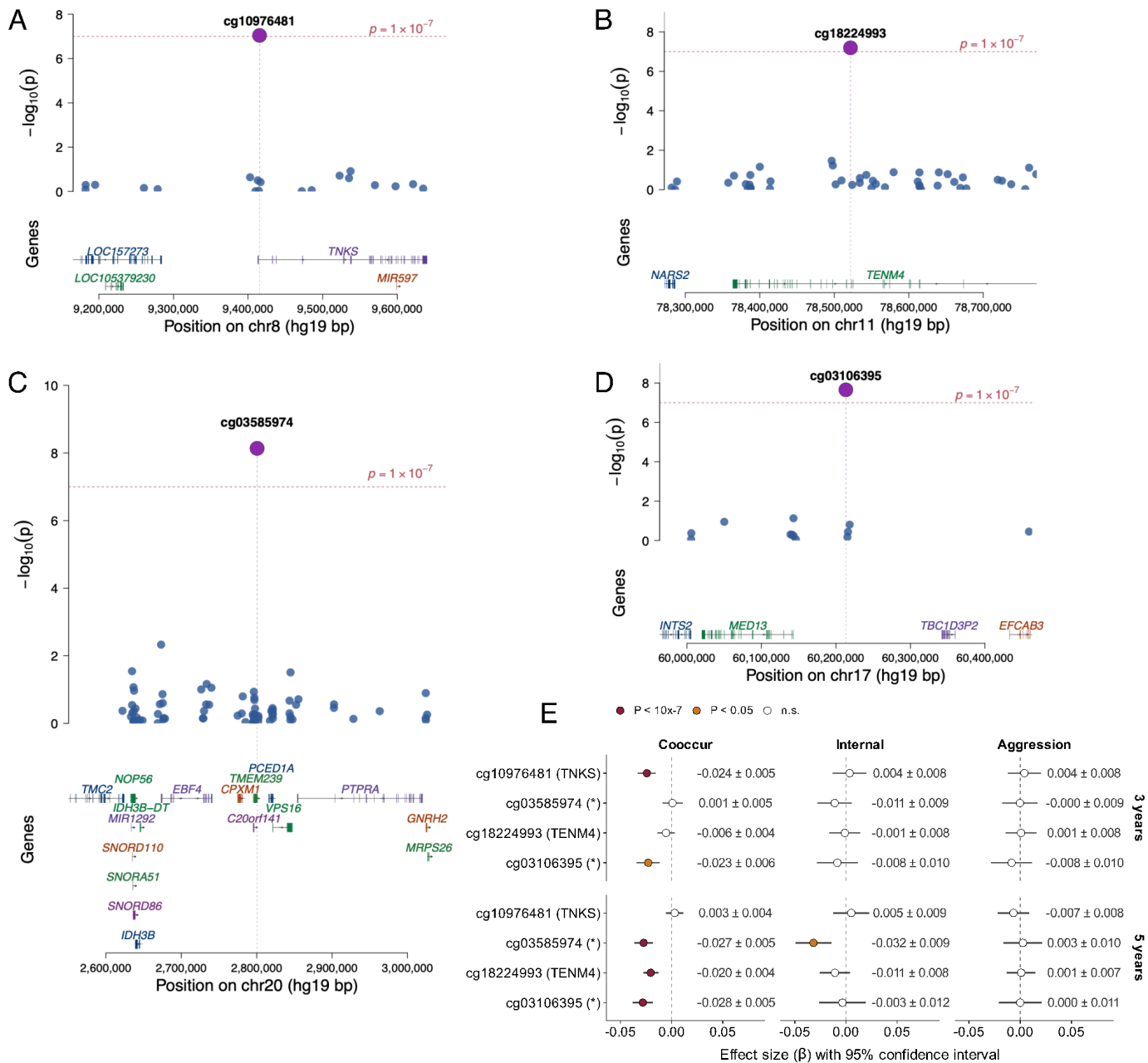


**Figure 3**

### Single-site cord-blood DNA methylation associations with childhood behavioral profiles.

(A) Regional methylation association plot for cg10976481, annotated to *TNKS*, associated with co-occurring problems at age three. (B–D) Regional methylation association plots for age-five co-occurring problems, showing cg18224993 in the *TENM4* locus, cg03585974 near *TMEM239*, and the intergenic CpG cg03106395, respectively. The dashed horizontal line indicates the epigenome-wide significance threshold. (E) Effect-size summary for the four differentially methylated CpGs across behavioral profiles and ages. Points show regression coefficients and horizontal lines show 95% confidence intervals for co-occurring, internalizing, and aggression profiles at ages three and five. Filled points indicate epigenome-

wide significant associations, colored points indicate nominal associations, and open points indicate non-significant associations.

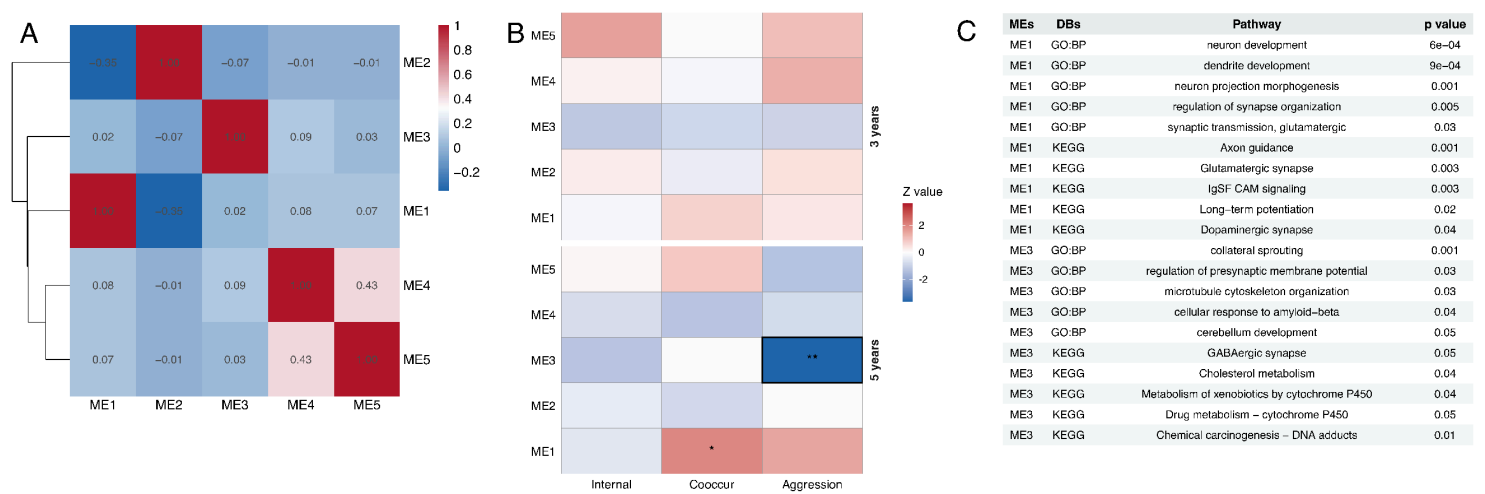


Figure 4

Co-methylation network analysis of cord-blood DNA methylation.

(A) Correlation heatmap of weighted co-methylation network analysis module eigengenes. Five modules were identified. ME1 and ME2 showed a negative eigengene correlation, whereas ME4 and ME5 showed a positive eigengene correlation, suggesting partially separable axes of methylation variation. (B) Associations between methylation module eigengenes and behavioral profiles at ages three and five. Heatmap colors represent regression Z values from logistic regression models adjusted for relevant covariates. Rows correspond to module eigengenes and columns to internalizing, co-occurring, and aggression profiles. Asterisks indicate nominally significant module–phenotype associations. (C) Selected nominally enriched pathways for modules associated with age-five behavioral profiles. The table shows biologically prioritized Gene Ontology Biological Process and KEGG pathways for ME1 and ME3, together with nominal enrichment P values. Pathways were selected to summarize neurodevelopmental, synaptic, neuronal-signaling, cytoskeletal, lipid-metabolic, and xenobiotic-response themes. These enrichments did not survive multiple-testing correction and should be interpreted as hypothesis-generating.

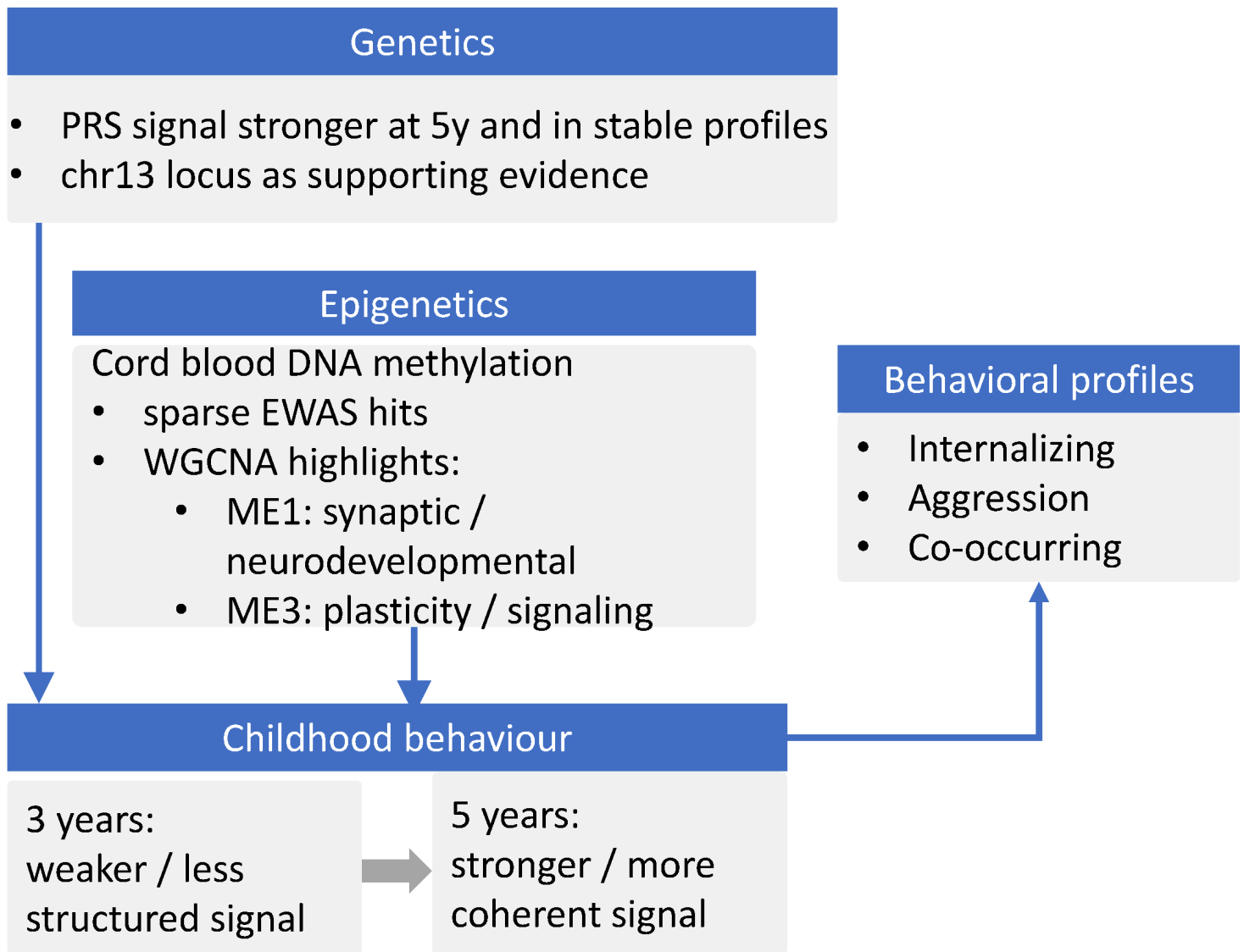


**Figure 5**

**Conceptual model summarizing developmental genetic and epigenetic correlates of early childhood behavioral profiles.**

The integrated model summarizes the main findings across molecular layers. Genetic analyses showed stronger polygenic score signals at age five and in stable behavioral profiles, with the chromosome 13q12 GWAS locus providing supporting locus-level evidence. Cord-blood DNA methylation analyses identified sparse single-site EWAS associations, whereas WGCNA highlighted methylation modules with nominally enriched neurodevelopmental, synaptic, plasticity-related, and signaling-related pathways. Across molecular layers, behavioral profiles at age three showed weaker and less structured signal, whereas age-five profiles showed stronger and more coherent molecular correlates. The model emphasizes internalizing, aggression, and co-occurring profiles as developmentally structured behavioral phenotypes rather than fixed diagnostic categories.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplTablesBekkhuss.xls](#)
- [SupplfiguresBekkhuss.pdf](#)