

Genome-wide analysis of social behaviour across social domains, reporters and developmental stages: a meta-regression approach

Received: 4 September 2025

Accepted: 14 July 2026

Published online: 21 September 2026



A list of authors and their affiliations appears at the end of the paper

Social behaviour is a heritable trait linked to well-being and mental health, yet it varies across social context, including social domains, informants and development, which may shape its underlying genomic architecture. Here, we conducted a genome-wide meta-regression meta-analysis of social behaviour from infancy to early adulthood, studying ~500,000 repeated measures in European-ancestry cohorts ($N = 73,321$). We modelled heterogeneity in single-variant associations across social domains (low prosocial behaviour and peer/social difficulties), multiple informants (parents, teachers and self-reports) and ages (2–29 years), identifying 6 associated loci, including variation within *CADM2* ($P = 2.51 \times 10^{-9}$). Single nucleotide polymorphism-based heritability ranged from 2% to 7%, and the genetic architecture comprised 4 factors reflecting differences in either social domains or informants. Polygenic scores ($N = 16,305$) accurately predicted contextual variation in independent cohorts, with some cross-ancestry transferability. Context-specific patterns of genetic correlations, particularly when estimating the developmental onset of associations, offer opportunities to disentangle shared trajectories linking social behaviour with later-life and mental health outcomes.

Human social behaviour encompasses a broad spectrum of interindividual interactions that change across development and vary across social situations^{1,2}. Prosocial behaviour involves voluntary actions that benefit others, such as helping, sharing and comforting. Early forms of prosocial behaviour can be observed from the age of 2 years³, and mastering increasingly complex levels of social competence during development shapes social roles in society². Social behaviour also includes peer and social difficulties, such as challenges to conform to developmental expectations, poor social skills, difficulties developing friendships and peer rejection^{4,5}. In particular, childhood peer difficulties have been associated with lower academic performance, self-esteem, social engagement, problem-solving skills and resilience^{6,7}. Thus, early social behaviour plays an important role in later-life friendships, well-being and (mental) health^{6–8}.

Twin studies have shown that both prosocial behaviour and peer and social problems (PSP) during childhood and adolescence

are heritable, with twin heritability (h^2_{twin}) estimates of 26–73% (refs. 9,10) and 24–71% (refs. 11–13), respectively. Corresponding single nucleotide polymorphism (SNP)-based heritability (h^2_{SNP}) estimates for PSP ranged from 2% to 27% (refs. 11,14). Yet, the genetic architecture underlying behaviour may vary by age and informant (for example, parent versus teacher reports)^{9,11,15,16}. Therefore, differences in genetic effects may also reflect variation in developmental processes driving social behaviour in different social situations. However, a systematic genomic investigation of the contextual factors contributing to heterogeneity in prosocial behaviour and PSP is still outstanding, in particular as large-scale genome-wide studies of social behaviour-related traits have exclusively been conducted in adult samples^{17–20}. In turn, matching contextual factors between discovery and target phenotypes may systematically affect the accuracy of derived polygenic scores (PGS), necessitating the calibration of polygenic analyses to improve power²¹, and have not yet been studied for social behaviour.

A full list of affiliations appears at the end of the paper. ✉ e-mail: beate.stpourcain@mpi.nl

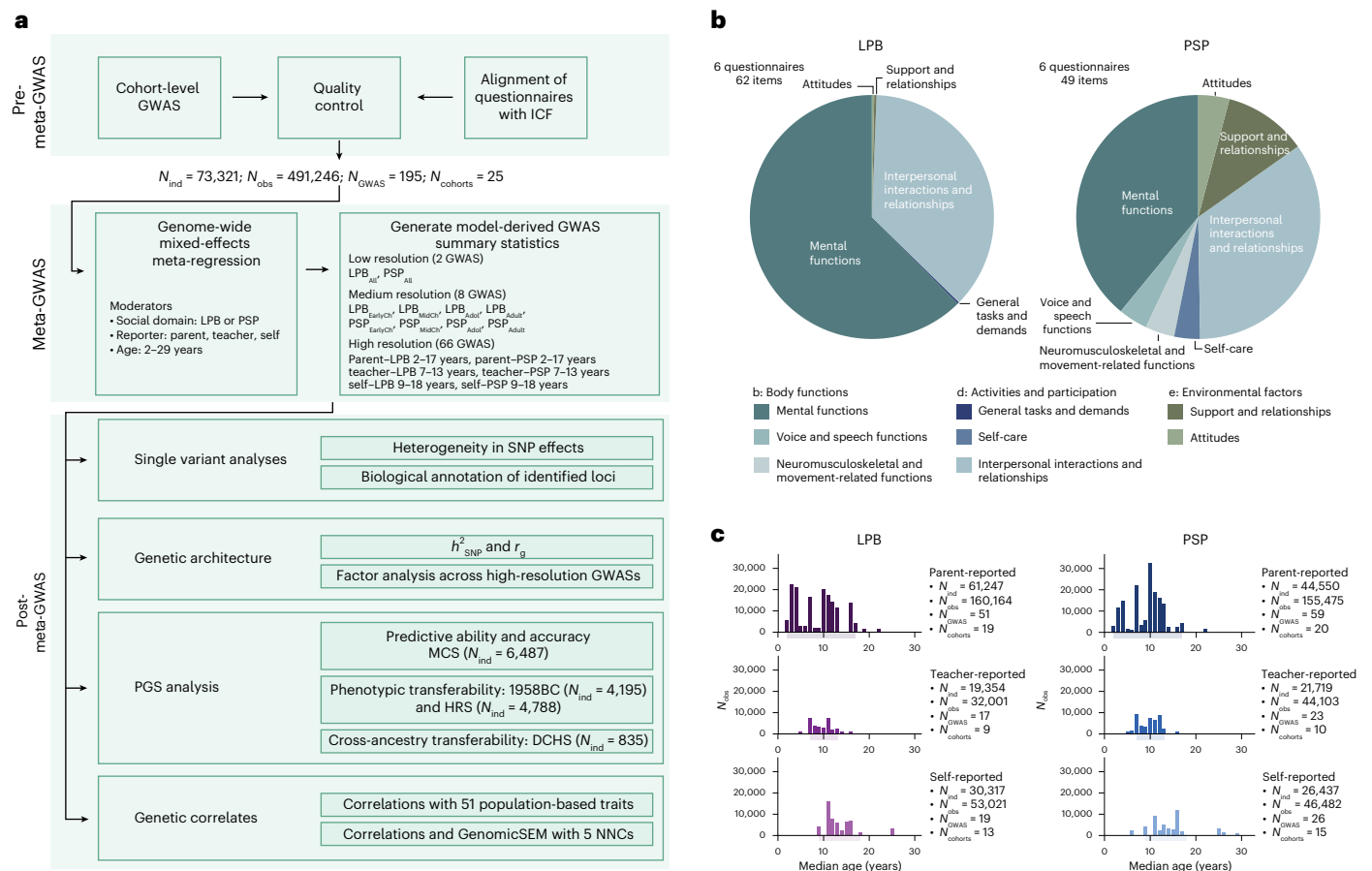


Fig. 1 | Social behaviour meta-GWAS design. a, Analysis design including a meta-regression across two social domains, three reporters and ages. Model-derived GWAS summary statistics refer to different social domains (LPB, low prosocial behavior; PSP, peer and social problems), informants (parent, teacher, self-report), ages (2–18 years) or developmental windows (EarlyCh, early childhood (ages 2–5 years); MidCh, mid-childhood (ages 6–11 years); Adol, adolescence (ages 12–17 years); Adult, young adulthood (ages 18–29)). **b**, Distribution of ICF codes

(codes b, body functions; d, activities and participation and e, environmental factors) per social domain. **c**, Distribution of individuals across age, per social domain and reporter, in addition to the number of observations (N_{obs}), the number of individuals (N_{ind}), the number of sets of GWAS statistics (N_{GWAS}) and number of cohorts ($N_{cohorts}$). 1958BC, 1958 birth cohort; DCHS, Drakenstein Child Health Study; HRS, Health and Retirement Study.

In this work, we present a genome-wide meta-analysis of social behaviour from infancy to young adulthood that explicitly accounts for and models variation in social context when estimating genetic associations. Adopting a versatile meta-regression framework²², we model heterogeneity in genetic effects across different social domains (capturing different aspects of social functioning), informants (capturing different social situations) and ages (capturing developmental change), all of which are included as moderators in our genetic association analysis. Here, we report SNP effects and describe the genetic architecture of social behaviour at different contextual (that is, moderator) levels. We subsequently construct PGS and demonstrate that their predictive ability and accuracy depend on matching contextual factors. We report genetic correlates of social behaviour, including population-based traits and neurodevelopmental and neuropsychiatric conditions (NNCs), with contextual resolution. Specifically, we proxy developmental trajectories of shared genomic dimensions, including the age of onset of genetic associations with social behaviour. Together, these analyses facilitate a context-sensitive study of genome-wide associations (GWAs) and provide insights into the genetic underpinnings of social behaviour.

Results

Modelling heterogeneity through genome-wide meta-regression analysis

We conducted a GWA study (GWAS) meta-analysis of social behaviour by integrating 491,246 repeated measures from 73,321 individuals of

European ancestry across 25 independent cohorts (effective sample size $N_{eff} = 122,239$; Fig. 1, Extended Data Table 1, Supplementary Table 1 and Supplementary Note 1), embedded within the Early Genetics and Lifecourse Epidemiology (EAGLE) Consortium²³. This effort resulted in 195 sets of GWAS summary statistics, for which we recorded information on social context, including social domains, informant and age (Fig. 1c and Supplementary Table 1). The majority of cohorts consisted of unrelated individuals; for family-based cohorts, relatedness was accounted for in cohort-level GWAS analyses (Methods).

Social behaviour was predominantly measured with standardized instruments (Supplementary Table 2), such as the Strengths and Difficulties Questionnaire (SDQ)⁴ or the Child Behaviour Checklist (CBCL)²⁴. To harmonize measures across instruments, we aligned individual questionnaire items to categories of the World Health Organization (WHO) International Classification of Functioning, Disability and Health²⁵ (ICF; Children and Youth version), a biopsychosocial framework and classification system for describing and organizing information on human functioning and disability (Fig. 1b, Extended Data Fig. 1, Supplementary Table 3, Supplementary Note 2 and Methods).

To model differences in genetic effects due to variation in social context, we implemented a genome-wide mixed-effects meta-regression approach, using the metafor R package²⁶. We meta-analysed the 195 GWAS estimates, modelling SNP effects as a function of contextual moderators (meta-GWAS; Fig. 1a). These included social domain (low prosocial behaviour (LPB; reference) and PSP), informant (parent

Table 1 | Independent GWA loci for social behaviour across social domains, reporters and ages

Lead SNP	Position (hg19)	Effect allele	EAF	β (s.e.)	Strongest association (P value)	GWAS where hit is reported	Nearest gene	Max pLI (gene)	Max CADD (gene)
rs148774891	Chr. 3: 86,059,243	A	1.29%	0.202 (0.034)	Self-reported LPB at 9 years (2.51×10^{-9})	Self-reported LPB at 9–14 years	<i>CADM2</i> <i>CADM2-AS1</i>	0.98 (<i>CADM2</i>)	3.14 (<i>CADM2</i>)
rs181817841	Chr. 8: 78,481,590	T	1.25%	0.154 (0.026)	Parent-reported LPB at 6 years (3.33×10^{-9})	Parent-reported LPB at 3–8 years	<i>RP11-38H17.1</i>	—	5.12 (<i>RP11-38H17.1</i>)
rs7490422	Chr. 13: 23,284,225	G	35.04%	−0.035 (0.006)	Parent-reported PSP at 6 years (5.37×10^{-9})	Parent-reported PSP at 4–8 years	<i>DDX39AP1</i>	—	1.68 (<i>DDX39AP1</i>)
rs77125329	Chr. 10: 78,297,483	T	1.24%	−0.150 (0.026)	Parent-reported LPB at 4 years (5.70×10^{-9})	Parent-reported LPB at 2–5 years and LPB _{EarlyCh}	<i>LRMDA</i>	0.04 (<i>LRMDA</i>)	3.81 (<i>LRMDA</i>)
rs538645	Chr. 11: 118,711,069	T	76.43%	−0.044 (0.008)	PSP _{EarlyCh} (1.41×10^{-8})	PSP _{EarlyCh}	<i>SETP16</i> <i>Y_RNA</i>	1.00 (<i>ARCNI</i>)	14.02 (<i>SETP16</i> and <i>Y_RNA</i>)
rs150685860	Chr. 20: 30,561,451	A	2.09%	−0.164 (0.030)	Parent-reported PSP at 2 years (2.72×10^{-8})	Parent-reported PSP at 2–3 years	<i>XKR7</i>	0.92 (<i>XKR7</i>)	6.35 (<i>XKR7</i>)

SNP identifier, position, effect allele, EAF, effect size (β), GWAS with the reported SNP and P value (two-sided), nearest gene and the highest pLI and Combined Annotation-Dependent Depletion (CADD) scores for the genes mapped with FUMA based on positional and eQTL mapping are shown. Estimates for all GWA signals are provided in Extended Data Table 2 and Supplementary Tables 5 and 6, based on mixed-effects meta-regression models. Estimates for FUMA mapping are provided in Supplementary Table 7.

(reference), teacher and self; hereafter referred to as reporter) and age (2–29 years; centred). For each SNP, fixed effects (θ parameters) captured systematic changes in the genetic effect (β) across these moderators. The fixed-effect intercept (θ_0) represents the SNP effect for parent-reported LPB at the grand mean age (10.14 years). We approximated the sampling variance, allowing for repeat observations within cohorts, and accounted for between-cohort and within-cohort heterogeneity using random effects (Methods).

From each SNP-based model, genome-wide, we derived meta-analysed SNP effect sizes at different moderator levels, facilitating the harmonization of summary statistics across diverse social contexts into interpretable patterns (Supplementary Table 4 and Methods). This approach resulted in 3 groups of GWAS summary statistics ($N_{\text{GWAS}} = 76$) with increasing contextual resolution: (1) low resolution ($N_{\text{GWAS}} = 2$: LPB_{All} and PSP_{All}), differentiating between social domains; (2) medium resolution ($N_{\text{GWAS}} = 8$: early-childhood LPB (LPB_{EarlyCh}), mid-childhood LPB (LPB_{MidCh}), adolescent LPB (LPB_{Adol}), adult LPB (LPB_{Adult}), PSP_{EarlyCh}, PSP_{MidCh}, PSP_{Adol} and PSP_{Adult}), differentiating between social domains across broad developmental stages (early childhood, 2–5 years; middle childhood, 6–11 years; adolescence, 12–17 years; and young adulthood, 18–29 years); and (3) high resolution ($N_{\text{GWAS}} = 66$), differentiating between social domains across different reporters and individual ages. High-resolution GWAS allowed for the estimation of systematic and dynamic variation in SNP effects across the genome, including age-related changes, as visualized in dynamic Manhattan plots (Supplementary Videos 1–6).

Across the 76 derived GWAS, we identified 6 independent GWA loci ($P < 5 \times 10^{-8}$) with 1 lead variant per locus (lead SNPs, [rs148774891](#), [rs181817841](#), [rs7490422](#), [rs77125329](#), [rs538645](#) and [rs150685860](#)), as confirmed in conditional and joint analyses using GCTA v.1.94 (ref. 27) software (Methods (COJO), Table 1, Extended Data Table 2, Extended Data Figs. 2 and 3, Supplementary Tables 5 and 6, Supplementary Fig. 1, Supplementary Note 3 and Supplementary Videos 1–6). All six loci showed considerable heterogeneity in genetic effects, with moderating effects (θ) attributable to differences in social domains, reporters or age (Extended Data Table 2, Supplementary Table 5 and Supplementary Note 3); none had yet been reported.

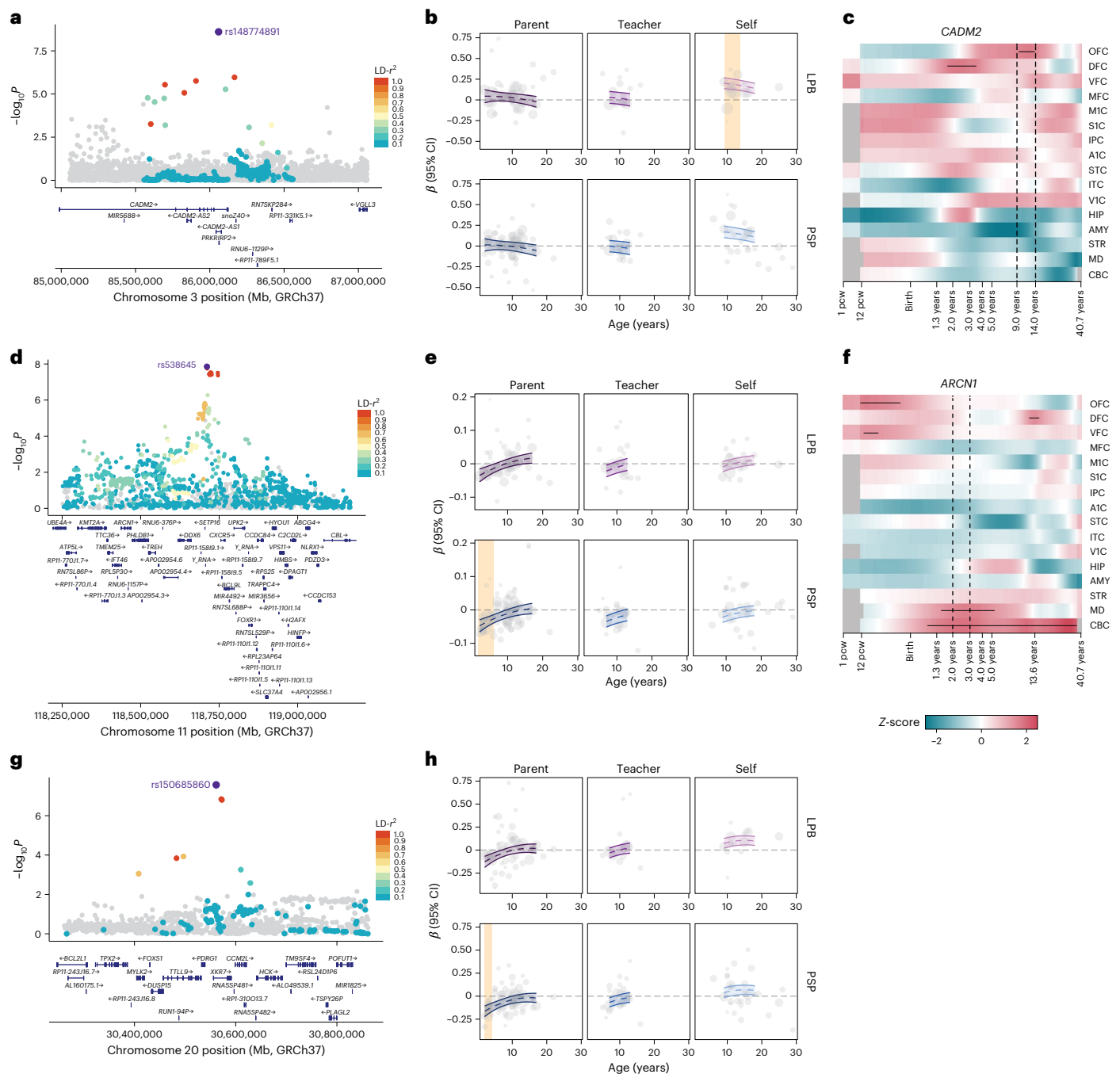
Biological annotation of single-variant signals

As part of post-meta-GWAS analyses (Fig. 1a, Supplementary Figs. 2–13, Supplementary Tables 7–10, Supplementary Note 4 and Methods), the 6 GWA loci were mapped to 77 genes (39 protein-coding genes) using FUMA²⁸ and MAGMA²⁹ software. Specifically, we conducted positional mapping and expression quantitative trait loci (eQTL) mapping (Supplementary Table 7), in-depth regulatory

variation (xQTL) screening for eQTL, isoform proportion quantitative trait loci (isoQTL) and methylation quantitative trait loci (mQTL) (Supplementary Fig. 2 and Supplementary Table 7), and chromatin interaction mapping (Supplementary Table 8). Here, we focus on three GWA loci ([rs148774891](#), [rs150685860](#) and [rs538645](#)) with probable biological impact (Fig. 2).

The strongest signal, [rs148774891](#), is a low-frequency variant (effect allele frequency (EAF) = 1.29%, A allele; Fig. 2a) located on chromosome 3p12.1 within an intron of the cell adhesion molecule 2 (*CADM2*) gene and *CADM2* antisense RNA 1 gene (*CADM2-AS1*). Each effect allele was associated with an increase in self-reported LPB at 9 years by 0.20 s.d. units (s.e. = 0.034, $P = 2.5 \times 10^{-9}$; Table 1), but the signal remained detectable until age 14. The *CADM2* gene, widely expressed in the brain, has a high score for probability loss-of-function intolerance (pLI = 0.98) and plays a vital role in synapse organization³⁰ (Supplementary Figs. 4 and 10). Matching the developmental window of the associated trait, *CADM2* overexpression peaks in the orbitofrontal cortex at 9–14 years (Fig. 2c and Supplementary Fig. 11), a region important for decision-making in social and reward-based contexts³¹.

The lead variant [rs538645](#) (EAF = 76.43%, T allele) on chromosome 11q23.3 was associated with a 0.044 s.d. decrease in early childhood PSP (ages 2–5 years; s.e. = 0.0077, $P = 1.4 \times 10^{-8}$; Fig. 2e and Table 1). The locus was mapped by position to the SET pseudogene 16 (*SETP16*, ENSG00000201535; *Y_RNA*; Supplementary Table 7), where variants within this locus have a Combined Annotation-Dependent Depletion score of 14.02, indicating potential deleteriousness. Furthermore, we identified [rs7104519](#), a variant in high linkage disequilibrium (LD) with [rs538645](#) (LD as measured by $r^2 = 1$), as an eQTL linked to a decrease in expression of archaia 1 (*ARCNI*) ($P_{\text{eQTL}} = 2.67 \times 10^{-7}$; Supplementary Table 7), which is a gene with a high pLI score (pLI = 1) expressed in the brain (Supplementary Fig. 8). Matching the developmental stage of the associated trait, *ARCNI* mRNA showed above-average expression in the cerebellar cortex and the mediodorsal nucleus of the thalamus across ages 1–5 years (Fig. 2f and Supplementary Fig. 11). In addition, xQTL mapping (Supplementary Fig. 2) revealed that [rs538645](#) is a direct mQTL previously associated with the demethylation of a CpG site within the Casitas B lineage lymphoma proto-oncogene (*CBL*), corresponding to an increase in *CBL* expression in cortex brain tissue (in relation to [rs538645](#)-T, $P_{\text{mQTL}} = 1.47 \times 10^{-17}$; Supplementary Table 7)³². A second variant in high LD with [rs538645](#) ([rs566848](#), LD- $r^2 = 0.97$) was identified as an isoQTL most strongly linked to decreased expression of the human RNA helicase *DDX6* gene (in relation to [rs538645](#)-T, $P_{\text{isoQTL}} = 7.43 \times 10^{-12}$) in the brain cortex³³, a gene with a high pLI score (pLI = 1) that was also found through chromatin interaction mapping (Supplementary Table 8).



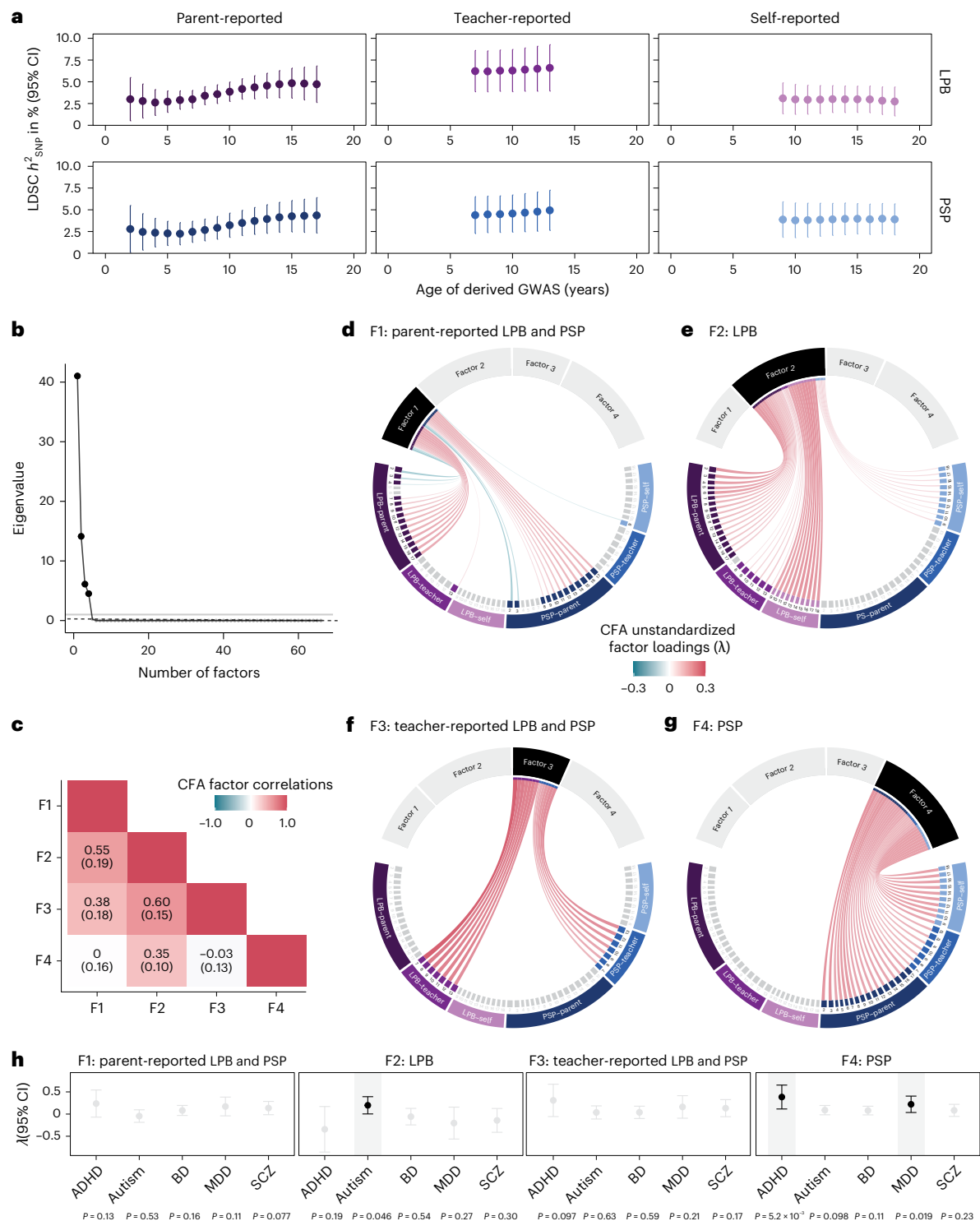


Fig. 3 | Genetic architecture of social behaviour. **a**, LDSC h^2_{SNP} estimates across social context (that is, social domains, reporters and age; 66 high-resolution social behaviour GWAS). Each dot represents an h^2_{SNP} point estimate, and error bars indicate the corresponding 95% CIs. The sample size of each GWAS is presented in Supplementary Table 4. **b**, Scree plot of genetic eigenvalues. Each dot indicates an eigenvalue. The grey line corresponds to an eigenvalue of 1. The dashed black line shows the scree line inferred from the data. **c, d–g**, Factor correlations and corresponding unstandardized factor loadings from a correlated 4-factor genetic CFA model of social behaviour. **c**, Factor correlation values are shown with the corresponding 95% CI in brackets. **(d–g)**, Unstandardized factor loadings are shown for Factor 1 (**d**), Factor 2 (**e**), Factor 3 (**f**) and Factor 4 (**g**), where the colour represents the magnitude of each factor loading. Detailed model information, including all factor

loadings, is available in Supplementary Table 14. **h**, Mapping NNCs onto the four-factor social behaviour model as shown via factor loadings (λ). For each NNC, the four-factor social behaviour CFA model shown in panels **c–g** was fitted separately. Points represent the estimated unstandardized factor loading of the mapping of each NNC onto the 4 latent social behaviour factors, and error bars indicate 95% CIs. Factor loadings with $P \leq 0.05$ are shown in black, and grey otherwise (2-sided Wald test). The sample size of each social behaviour GWAS is presented in Supplementary Table 4. The sample sizes for the 5 NNCs are as follows: ADHD ($N_{\text{cases}} = 38,067$, $N_{\text{controls}} = 185,088$), autism ($N_{\text{cases}} = 18,381$, $N_{\text{controls}} = 27,969$), BD ($N_{\text{cases}} = 27,196$, $N_{\text{controls}} = 43,792$), MDD ($N_{\text{cases}} = 26,778$, $N_{\text{controls}} = 51,857$) and schizophrenia (SCZ; $N_{\text{cases}} = 53,386$, $N_{\text{controls}} = 77,258$). **d–h**, Note that unstandardized factor loadings reflect relationships at the level of the captured h^2_{SNP} .

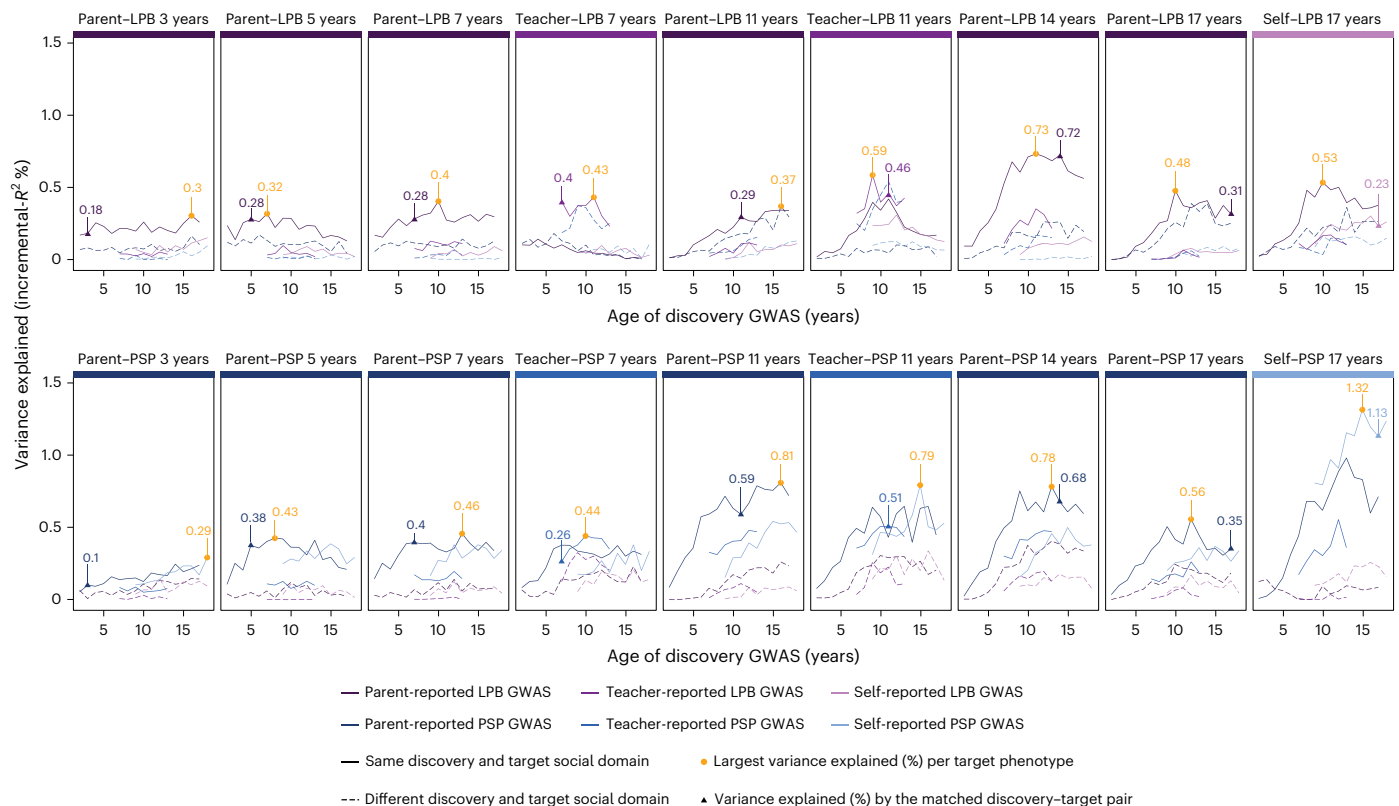


Fig. 4 | Polygenic predictability of social behaviour. Social measures in the MCS (target) were predicted with PGS from 66 high-resolution social behaviour GWAS, stratified by social domain, reporter and age (discovery). For each target phenotype (plot), the explained variance (incremental- R^2 percentage, y axis) is shown against the predicted age (in years) of the discovery GWAS. Blue lines indicate PSP discovery GWAS based on parent (darker), teacher (medium) and self (lightest) reports. Purple lines indicate LPB discovery GWAS based on parent

(darker), teacher (medium) and self (lightest) reports. Lines are solid if the target and discovery social domains match (for example, parent-reported LPB GWAS predicting parent-LPB at 3 years in the MCS) and dashed otherwise. For each target phenotype, the variance explained by a matching PGS, aligned by social domain, reporter and age, is highlighted with a triangle. For each target phenotype of the MCS, the largest variance explained by any discovery PGS is shown in orange.

Lastly, the low-frequency variant [rs150685860](#) (EAF = 2.09%, A allele) resides within an intron of the XK-related 7 gene (*XKR7*) on chromosome 20q11.21 (Fig. 2d), a gene with a high pLI score (pLI = 0.92) that is involved in apoptotic processes. *XKR7* is widely expressed in the cerebellum, but cortical and subcortical expression was low (Supplementary Fig. 9). Each effect allele was associated with a 0.16 s.d. decrease in parent-reported PSP at 2 years (s.e. = 0.029, $P = 2.7 \times 10^{-8}$; Table 1), indicating a rare protective effect limited to ages 2–3 years.

The genetic architecture of social behaviour is multidimensional

Social behaviour was heritable and GWAS were overall sufficiently powered, although h^2_{SNP} estimates were small, ranging from 2% to 7% (LD score regression (LDSC)³⁴; Fig. 3a and Supplementary Table 11). At low resolution, h^2_{SNP} was estimated at 3.43% (s.e. = 0.48%, Z-score = 7.15) for LPB_{All} and 3.11% (s.e. = 0.55%, Z-score = 5.65) for PSP_{All}. When allowing for differences in social domains and broad developmental stages (medium-resolution GWAS), the range of h^2_{SNP} estimates increased to 2.25% (s.e. = 2.34%, Z-score = 0.96, LPB_{Adult}) and 3.93% (s.e. = 0.62%, Z-score = 6.34, LPB_{Adol}). Further stratification by social domain, reporter and age (high-resolution GWAS) yielded an even broader range of h^2_{SNP} estimates from 2.15% (s.e. = 0.82%, Z-score = 2.62, parent-reported PSP at 4 years) to 6.67% (s.e. = 1.34%, Z-score = 4.98, teacher-reported LPB at 13 years), although 95% confidence bands were overlapping.

The range of genetic correlations, estimated with LDSC³⁵, also broadened with increasing GWAS resolution (Supplementary Figs. 14 and 15). Low-resolution GWAS (LPB_{All} and PSP_{All}) were moderately interrelated (genetic correlation estimate (r_g) = 0.58, s.e. = 0.096;

Supplementary Fig. 14a). At medium resolution, patterns of high genetic similarity emerged (Supplementary Fig. 14b). At high resolution, analyses revealed increasingly complex genetic interrelationships, highlighting the need for a structural model to describe the genetic architecture underlying social behaviour across social domains, reporters and developmental stages (Supplementary Fig. 15).

We dissected the genetic covariance across high-resolution GWAS ($N_{\text{GWAS}} = 66$; Supplementary Note 5) using exploratory factor analysis and confirmatory factor analysis (CFA), adopting a data-driven approach (Fig. 3b–g and Methods)³⁶. A 4-factor CFA model with correlated factors, implemented in Genomic Structural Equation Modelling (GenomicSEM)³⁷, fitted the data best (comparative fit index (CFI) = 0.98, standardized root mean square residual (SRMR) = 0.05; Fig. 3c–g, Supplementary Fig. 16, Supplementary Tables 12–14 and Supplementary Note 5). The 4 factors explained genetic variation as follows: factor 1 (F1) primarily reflected parent-reported LPB and PSP (largest factor loading: $\lambda_{F1} = 0.14$, s.e. = 0.033, parent-reported LPB at 14 years; Fig. 3d); factor 2 (F2) reflected parent-, teacher- and self-reported LPB (largest factor loading: $\lambda_{F2} = 0.18$, s.e. = 0.017, self-reported LPB at 15 years; Fig. 3e); factor 3 (F3) captured teacher-reported LPB and PSP (largest factor loading: $\lambda_{F3} = 0.27$, s.e. = 0.034, teacher-reported LPB at 7 years; Fig. 3f); and factor 4 (F4) captured parent-, teacher- and self-reported PSP (largest factor loading: $\lambda_{F4} = 0.17$, s.e. = 0.019, parent-reported PSP at 7 years; Fig. 3g). Thus, the identified factor structure aligned with either differences in social domains (LPB: F2; PSP: F4) or informants (parent: F1; teacher: F3). Factor correlations ranged from $r_g = -0.03$ (s.e. = 0.13) to $r_g = 0.60$ (s.e. = 0.15), with the

strongest correlations between F2 and F3 ($r_g = 0.60$, s.e. = 0.15) and F1 and F2 ($r_g = 0.55$, s.e. = 0.19; Fig. 3c). Notably, F1 captured an inverse association between parent-reported PSP in early childhood and LPB in middle and late childhood, consistent with bivariate correlations (for example, PSP at 2 years and LPB at 13 years, $r_g = -0.71$, s.e. = 0.22; Supplementary Fig. 15) and with qualitative developmental changes in social behaviour.

PGS predict contextual differences in social behaviour

We validated PGS for social behaviour in 4 European- and African-ancestry cohorts (number of individuals (N_{ind}) = 16,305, number of observations (N_{obs}) = 115,254; Supplementary Notes 6 and 7).

Studying children of European ancestry from the Millennium Cohort Study (MCS)³⁸, a cohort with extensive LPB and PSP phenotype data repeatedly assessed during childhood and adolescence by different informants using the SDQ⁴ ($N_{\text{ind}} = 6,487$, $N_{\text{obs}} = 96,259$; Supplementary Table 15), we evaluated the predictive ability and predictive accuracy of social behaviour PGS.

First, we confirmed the predictive ability of PGS by matching the contextual information of each discovery social behaviour GWAS to the corresponding social behaviour target in the MCS (18 LPB and PSP measures, ages 3–17 years, parent-, teacher- and self-reported; Supplementary Table 15). Matching discovery–target pairs yielded consistently positive effects (incremental $R^2 = 0.18$ – 1.13% ; Extended Data Fig. 4 and Supplementary Table 16), with all associations, except for parent-reported PSP at 3 years, passing the multiple-testing threshold ($P < 0.0037$; Methods).

Second, we evaluated the predictive accuracy of PGS as the percentage of associations with larger effects (in incremental R^2) for phenotypically matching discovery–target pairs than non-matching pairs (Supplementary Tables 17 and 18 and Supplementary Note 7), for all associations passing the multiple-testing threshold. For domain-matched PGS signals, the predictive accuracy was high, especially for the PSP social domain (PSP = 81.6%, LPB = 45.6%; Supplementary Table 17). For reporter-matched PGS, the predictive accuracy was highest for parent-reported (65.53%) and teacher-reported (66.07%) outcomes, followed by self-reports (38.8%). For both domain- and reporter-matched PGS, the predictive accuracy increased even further (parent-report outcomes, 88.6%; teacher-report outcomes, 71.4%; self-report outcomes, 50%), although the predictive accuracy for self-reported LPB using self-report PGS was poor, and the outcome was better captured by parent-reported PGS (Supplementary Fig. 17 and Supplementary Table 18). Self-reported prosocial behaviour measures are prone to socially desirable responding³⁹ and this bias may affect genomic analyses, especially when using single-cohort self-report measures.

Third, comparing the largest PGS signal (in incremental R^2) with the fully matching PGS, we observed, for associations passing the multiple-testing threshold, 89% domain-matched, 78% reporter-matched and 78% age-matched (within a 5-year age window) signals capturing contextual information of the target phenotype, with similar magnitude (Fig. 4, Supplementary Table 17 and Supplementary Note 7). Thus, while matching discovery and target characteristics increased predictability (incremental $R^2 \leq 1.32\%$; Fig. 4, Supplementary Figs. 18 and 19, and Supplementary Table 16), meta-regression-derived PGS also predicted contextual target information, if unmatched, except for self-report LPB outcomes.

Subsequently, we corroborated the phenotypic transferability of social behaviour PGS to broadly defined social measures assessed with non-standardized instruments during early and late adulthood, studying 2 further European-ancestry samples (incremental $R^2 \leq 0.22\%$, total $N_{\text{ind}} = 8,983$; Supplementary Tables 15, 19 and 20 and Supplementary Note 7).

Lastly, we predicted SDQ and CBCL measures in an African-ancestry cohort ($N_{\text{ind}} = 835$; Supplementary Tables 15 and 21 and Supplementary Note 7). Matching social domains and developmental stages, social behaviour GWAS PGS were associated with parent-reported LPB at ages 7 and 8 years (incremental $R^2 \leq 0.95\%$), highlighting the cross-ancestry transferability of LPB PGS.

Genetic correlates of social behaviour reveal contextual association patterns

With PGS systematically capturing context-dependent phenotypic differences, we also investigated genetic correlates of social behaviour, allowing for contextual variation. In particular, we studied overlap with high-resolution GWAS that differentiate between social domains, informants and ages to identify contextual correlation patterns, with developmental links approximating the onset of phenotypic associations with social behaviour.

First, we examined genetic correlations across a spectrum of population-based traits ($N_{\text{traits}} = 51$), including externalizing, internalizing, social, well-being, personality, cognition and educational attainment (EA), and language, in addition to anthropometric and birth measures, using LDSC software³⁴ (multiple-testing threshold of $P < 8.89 \times 10^{-4}$; Figs. 5 and 6, Extended Data Figs. 5 and 6, Supplementary Tables 22–24 and Methods).

The largest and strongest correlates of social behaviour were behavioural outcomes (Fig. 5 and Supplementary Table 23). Genetic correlations with (near) adult externalizing traits were positive and distinctly larger for PSP compared with LPB, with non-overlapping 95% confidence bands (for example, aggression: $\text{PSP}_{\text{All}} r_g = 0.96$, s.e. = 0.10; $\text{LPB}_{\text{All}} r_g = 0.51$, s.e. = 0.09), indicating domain-specific effects. For non-substance-use-related externalizing behaviour, the predicted onset of PSP associations preceded LPB associations, especially for parent-reported PSP (Fig. 6 and Extended Data Fig. 7). For (near) adult internalizing traits (for example, internalizing behaviour: $\text{PSP}_{\text{All}} r_g = 0.9$, s.e. = 0.14; loneliness: $\text{PSP}_{\text{All}} r_g = 0.6$, s.e. = 0.08; Extended Data Fig. 8), many lack-of-well-being traits, substance-use-related traits and neuroticism, we observed moderate-to-strong PSP-specific correlations that were predicted to emerge in early childhood, for example based on parent-reported GWAS (Figs. 5 and 6, Extended Data Figs. 5 and 6, and Supplementary Table 24). Genetic correlations with EA and cognition were modest, inverse and PSP-specific (for example, EA: $\text{PSP}_{\text{All}} r_g = -0.26$, s.e. = 0.04). In contrast, for social outcomes and agreeableness, the predicted onset of LPB associations preceded that of PSP associations, especially in parent-reported GWAS (Fig. 6, Extended Data Figs. 5 and 6, and Supplementary Table 24). Genetic correlations consistently involved LPB, with low-to-moderate effects that were either similar for both LPB and PSP (for example, family relationship dissatisfaction: $\text{LPB}_{\text{All}} r_g = 0.38$, s.e. = 0.08; $\text{PSP}_{\text{All}} r_g = 0.30$, s.e. = 0.09) or LPB-specific (for example, low frequency of family and/or friends visits: $\text{LPB}_{\text{All}} r_g = 0.28$, s.e. = 0.06).

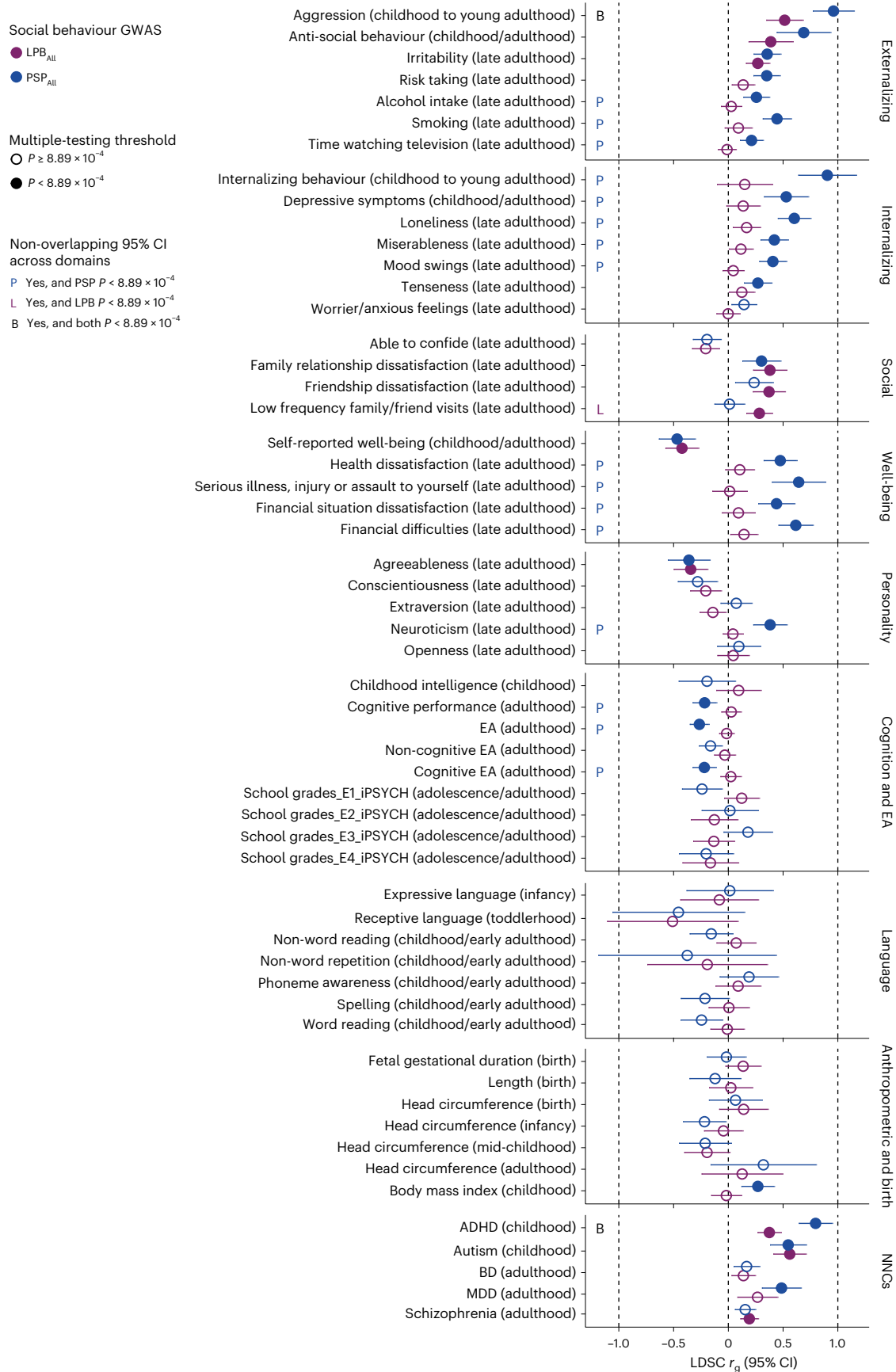
We next investigated the genetic correlation profile for five clinically defined NNCs (Figs. 3h, 5 and 7 and Supplementary Tables 24–29): attention deficit hyperactivity disorder (ADHD), autism, bipolar

Fig. 5 | Genetic correlations of low-resolution social behaviour with population-based traits and NNCs. LDSC genetic correlation estimates (r_g) with LPB_{All} and PSP_{All} GWAS are shown along with the corresponding 95% CIs. Filled dots indicate genetic correlations passing the multiple-testing threshold ($P < 8.89 \times 10^{-4}$, 2-sided Wald tests). r_g estimates, s.e. and P values are provided in Supplementary Table 23. Differences between PSP and LPB genetic correlations, consistent with trait-specific effects indicated by non-overlapping

95% confidence bands, are shown for PSP correlations (blue P), LPB correlations (purple L) or both (black B), once they pass the multiple-testing threshold. The maximum GWAS sample sizes for the social behaviour GWAS are as follows: $\text{LPB}_{\text{All}} N = 114,279$; $\text{PSP}_{\text{All}} N = 96,597$. Sample sizes for population-based traits and NNCs are presented in Supplementary Table 22. Note that GWAS for school grades refer to four principal components (E1–E4).

disorder (BD), major depressive disorder (MDD) and schizophrenia. For ADHD, both PSP- and LPB-specific associations were observed with larger correlations for PSP ($PSP_{All}r_g = 0.80$, s.e. = 0.078; $LPB_{All}r_g = 0.37$, s.e. = 0.055), particularly with parent- and teacher-reported GWAS

(Figs. 5 and 7a). MDD was moderately and positively associated with PSP ($PSP_{All}r_g = 0.49$, s.e. = 0.091) throughout, and only related to LPB for parent-reported GWAS at 13–17 years (Fig. 7a). For both ADHD and MDD, the predicted onset of PSP correlations in early childhood and



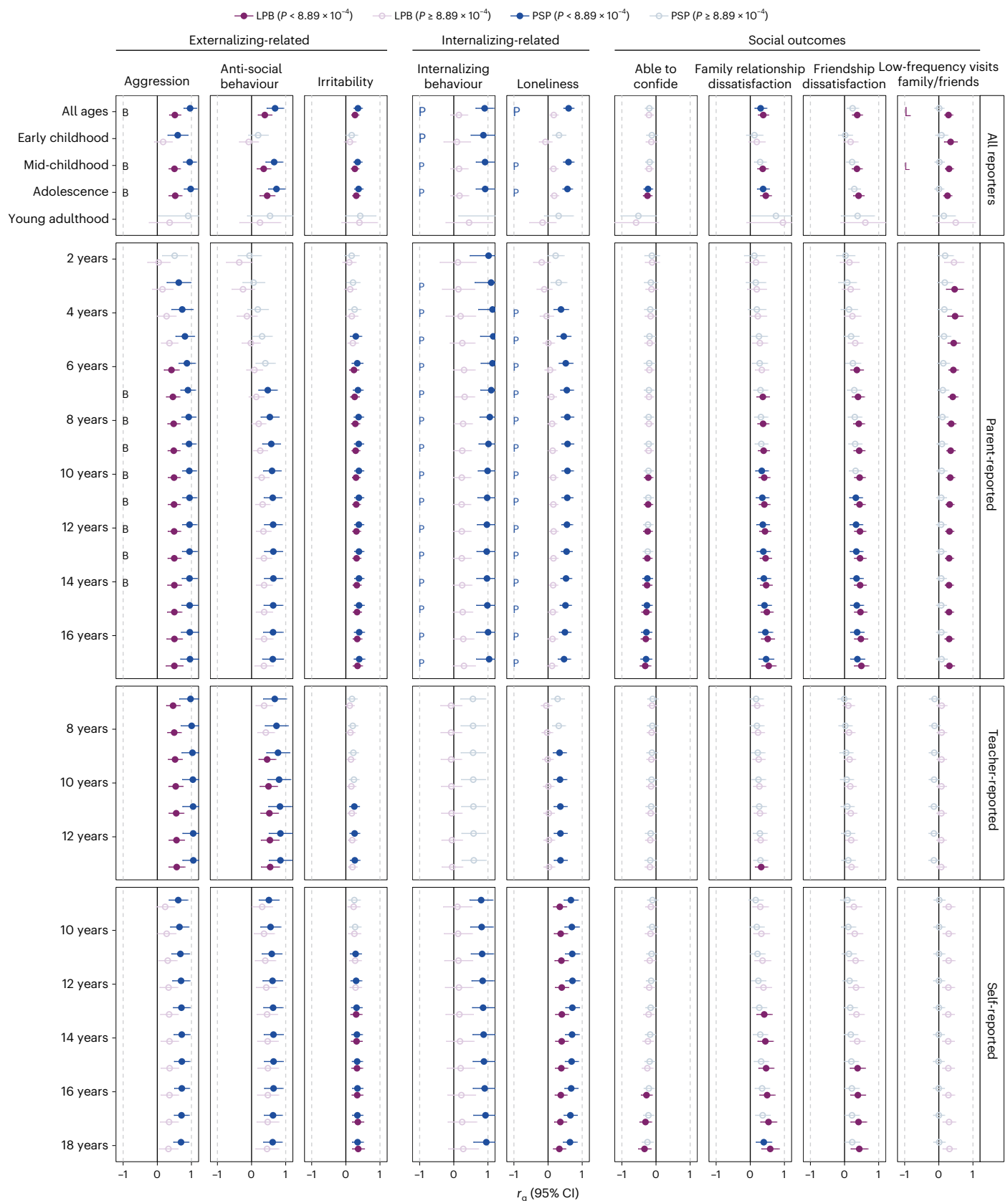


Fig. 6 | Contextual genetic correlations of social behaviour with (near) adult externalizing, internalizing and social outcomes. LDSC genetic correlation estimates (r_g) with low-resolution, medium-resolution and high-resolution LPB and PSP GWAS are shown along with the corresponding 95% CIs, with dashed vertical lines indicating the lower (-1) and upper (1) bounds of r_g . Filled dots indicate genetic correlations passing the multiple-testing threshold ($P < 8.89 \times 10^{-4}$, 2-sided Wald tests). r_g estimates, s.e. and P values are provided in

Supplementary Table 24. Differences between PSP and LPB genetic correlations at selected moderator levels, consistent with trait-specific effects indicated by non-overlapping 95% confidence bands, are shown for PSP correlations (blue P), LPB correlations (purple L) or both (black B) once they pass the multiple-testing threshold. GWAS sample sizes for social behaviour are presented in Supplementary Table 4. GWAS sample sizes for externalizing, internalizing and social outcomes are presented in Supplementary Table 22.

mid-childhood was earlier than for LPB correlations, especially for parent-reported GWAS (Fig. 7a), and both were linked to the social behaviour PSP factor (F4; Fig. 3h). Autism was moderately correlated with both social domains (LPB $r_g = 0.56$, s.e. = 0.08; PSP $r_g = 0.55$, s.e. = 0.08; Fig. 5) across all reporters, with early-onset associations, especially as predicted by parent-reported GWAS (Fig. 7a). For self-reported LPB, genetic correlations with autism were consistent with an r_g of 1 (for example, at 15 years; LPB $r_g = 0.80$, s.e. = 0.15; Fig. 7a), indicating a strong genetic overlap between population-based and clinical social dimensions⁴⁰, consistent with genetic links between autism and the social behaviour LPB factor (F2; Fig. 3h). Schizophrenia showed a modest genetic overlap with LPB (LPB $r_g = 0.19$, s.e. = 0.042), most apparent for parent- and teacher-reported GWAS (Fig. 7a), in line with our previous work⁴¹. Several parent-reported PSP GWAS (ages 7–17 years) also shared genetic links with schizophrenia, but the predicted onset of LPB correlations preceded PSP correlations (if present) in mid-childhood (Fig. 7a). Schizophrenia was not related to a social behaviour factor (Fig. 3h). In line with previous studies⁴¹, there was little evidence for an association between social behaviour and BD.

To triangulate findings, we also fitted a structural model across NNCs and medium-resolution social behaviour GWAS (Supplementary Tables 30–32), leveraging the larger h^2_{SNP} of NNCs for factor identification. Using GenomicSEM, we identified a correlated 4-factor CFA model with a good model fit (CFI = 0.99, SRMR = 0.069; Fig. 7b, Supplementary Fig. 20, Supplementary Note 8 and Methods). Two intercorrelated factors ($r = 0.59$, s.e. = 0.13) linked NNCs and social behaviour: F_{ADHD} (most strongly related to ADHD, factor loading $\lambda = 0.37$, s.e. = 0.08) explaining shared genetic variation underlying ADHD, MDD and PSP across development (Fig. 7b), and F_{Autism} (most strongly related to autism, factor loading $\lambda = 0.34$, s.e. = 0.06), capturing genetic variation shared between autism, LPB across development and, weakly, schizophrenia. Both F_{ADHD} and F_{Autism} were related to EA once mapped onto the factor structure (Supplementary Fig. 21 and Supplementary Table 33), consistent with socio-economic contributions.

Comparison of meta-regression with conventional meta-analytic approaches

We compared our meta-regression-based findings with those from a conventional meta-analysis approach, such as multivariate N-GWAMA⁴² that assumes no contextual heterogeneity (Supplementary Note 9 and Supplementary Fig. 22). Notably, the magnitude and strength of h^2_{SNP} estimates decreased when adopting a conventional approach (Supplementary Table 34). Still, LDSC intercepts, which capture the mean inflation due to confounding bias³⁴, were below one for all meta-regression GWAS (Supplementary Table 11), suggesting an (overly) conservative approach when accounting for sample overlap. Genetic correlations across low-resolution social domains derived with either method were strong (LPB GWAS, $r_g = 0.75$, s.e. = 0.04; PSP GWAS, $r_g = 0.89$, s.e. = 0.04; Supplementary Fig. 23 and Supplementary Table 34) and genetic correlation patterns with population-based traits and NNCs were highly consistent (Supplementary Fig. 24 and Supplementary Table 35), indicating that the presented results are robust to methodological choices.

Discussion

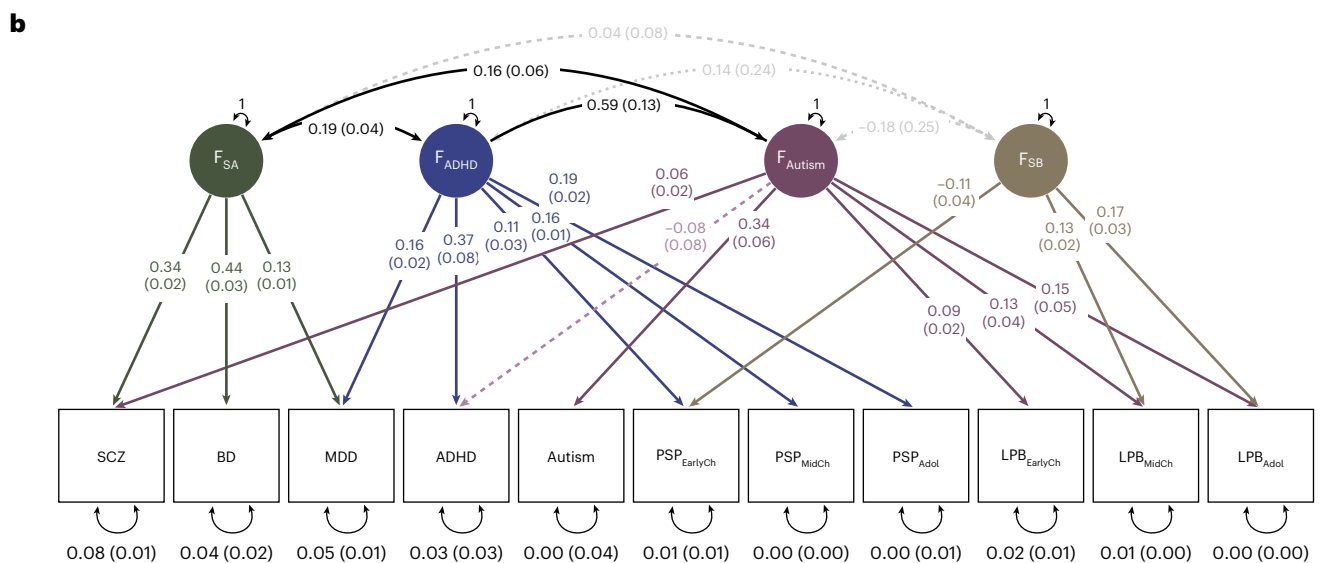
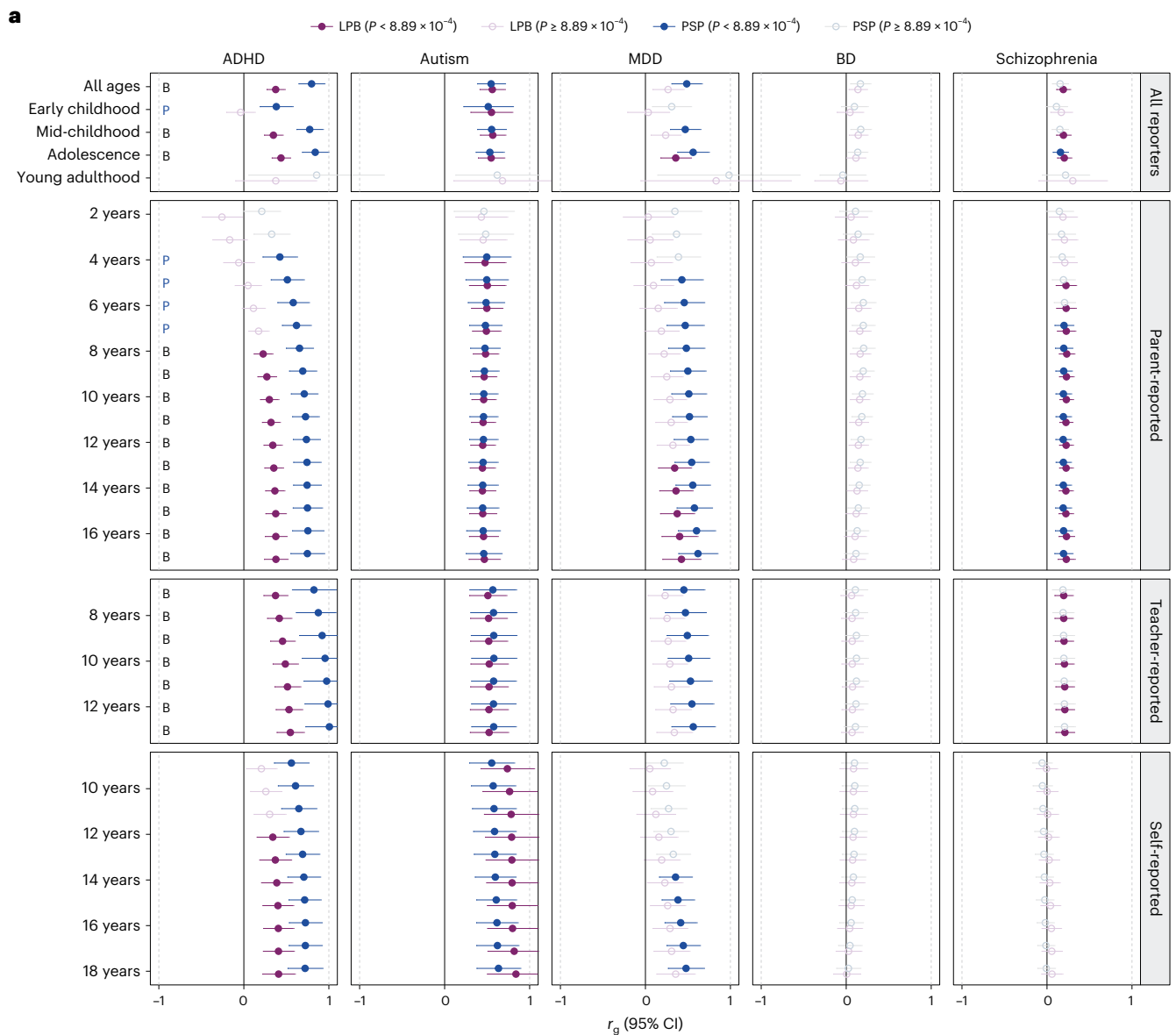
Adopting a genome meta-regression approach ($N_{\text{eff}} = 122,239$, $N_{\text{ind}} = 73,321$, 491,246 repeat observations) and following up findings in European- and African-ancestry cohorts ($N_{\text{ind}} = 16,305$), this study demonstrates that social behaviour is heterogeneous with systematic changes in genetic effects across social domains, informants and age. Here, we (1) identified GWAS signals with potential biological relevance, (2) characterized the multidimensional genetic architecture of social behaviour, (3) generated and validated PGS predicting the context of social behavioural difficulties, (4) uncovered genetic correlates of social behaviour with contextual association patterns, and (5) gained insight into the shared genetic architecture with population-based traits and NNCs, including the approximated age of association onset.

Our findings confirm that social behaviour is a highly context-dependent trait, with variation in social context being captured by both differences in single-variant effects and the genetic architecture (Box 1). We report six GWA loci associated with social behaviour, all of which had either domain-varying, age-varying and/or reporter-varying genetic effects. Two loci (**rs148774891** and **rs150685860**) were mapped to genes with a high mutation intolerance ($pLI > 0.9$), including *CADM2* and *XKR7*, respectively. **rs148774891** captured a 0.2 s.d. increase in self-reported childhood LPB per effect allele and was mapped to *CADM2*, a gene with development-specific expression patterns in the orbitofrontal cortex. Genetic variation within *CADM2* (unrelated to **rs148774891**, $LD-r^2 < 0.1$) has previously been linked to a wide spectrum of mental health-related traits⁴³, including autism^{44,45}, but also risk-taking behaviour⁴⁶, smoking and alcohol consumption⁴⁷, externalizing behaviour⁴⁸, EA⁴⁹, general cognitive ability⁵⁰, and the frequency of friends and family visits⁴³. *XKR7* has not yet been implicated in a known GWAS, but the rare protective effect at **rs150685860**, associated with a 0.16 s.d. decrease in early childhood parent-reported PSP per effect allele, suggests possible resistance mechanisms with translational potential. Another locus (**rs538645**) was identified as a direct mQTL for *CBL* and linked to *DDX6* via isoQTL mapping and to *ARCNI* via eQTL mapping. Rare germline mutations within *CBL* are known to cause a Noonan syndrome-like developmental disorder⁵¹, including developmental delay and cognitive impairment, while rare de novo mutations in *DDX6* are involved in a neurodevelopmental syndrome with intellectual disability⁵². The *ARCNI* gene, mapped to **rs538645**, is overexpressed in a development-dependent manner in subcortical brain regions that play a role in regulating social behaviour⁵³, while rare germline mutations within *ARCNI* cause *ARCNI*-related syndrome, often implicating developmental delays⁵⁴.

The SNP-based heritability of social behaviour was low ($h^2_{\text{SNP}} = 2\%$), but comparable to adult social traits^{17,19,20,55}, and most social behaviour GWAS were sufficiently powered (mean h^2_{SNP} Z-score = 4.37; range, 0.96–7.15), except for medium-resolution adult GWAS and very early childhood GWAS (h^2_{SNP} Z-score ≤ 3). The identification of four correlated genetic factors using a genetic CFA model underlines the heterogeneous nature of social behaviour¹, disentangling differences in informant characteristics from differences in social domains, of which only the latter were related to NNCs.

Fig. 7 | Genetic architecture across social behaviour and NNCs. a, Contextual genetic correlations of social behaviour with NNCs. LDSC genetic correlation estimates (r_g) with low-resolution, medium-resolution and high-resolution LPB and PSP GWAS are shown with their 95% CIs. Filled dots indicate genetic correlations passing the multiple-testing threshold ($P < 8.89 \times 10^{-4}$, 2-sided Wald tests). r_g , s.e. and P values are provided in Supplementary Table 24. Differences between PSP and LPB genetic correlations at selected moderator levels, consistent with trait-specific effects indicated by non-overlapping 95% confidence bands, are shown for PSP correlations (blue P), LPB correlations (purple L) or both (black B), once they pass the multiple-testing threshold. Sample sizes for social behaviour GWAS are presented in Supplementary Table 4. Sample sizes for NNCs GWAS are presented in Supplementary Table 22.

b, Genetic CFA across medium-resolution social behaviour and NNCs. The CFA model showed excellent model fit (CFI = 0.99, SRMR = 0.069). Observed measures are represented by squares (here GWAS statistics) and latent variables by circles. Coloured, single-headed arrows define unstandardized factor loadings (shown with their corresponding s.e.). Factor loadings with $P \leq 0.05$ (2-sided) are solid, and dashed otherwise. Double-headed arrows define factor correlations (shown with their corresponding s.e.). Factor correlations are shown with a black solid line if $P \leq 0.05$ and with a dashed grey line otherwise. Note that unstandardized factor loadings reflect relationships at the level of the captured h^2_{SNP} . Analyses were restricted to social behaviour GWAS statistics with sufficient power (h^2_{SNP} Z-score > 3). F_{ADHD} , ADHD genetic factor; F_{Autism} , autism genetic factor; F_{SA} , schizoaffective genetic factor; F_{SB} , social behaviour genetic factor.



BOX 1

Cautionary notes

Social behaviour reflects a broad behavioural continuum. While social behavioural difficulties are often linked to NNCs, variation in social behaviour can also be observed in typically developing children. We therefore provide guidance to help readers interpret the complex relationships underlying our findings.

Single-variant genetic effects reported in this study refer to the estimated regression coefficient of each SNP and do not imply causality. These genetic associations capture, for each variant, only a very small proportion of the phenotypic variance in experiencing social difficulties across the life course. Thus, the presence or absence of a variant cannot predict an individual's social behaviour. Nevertheless, identifying associated variants remains important for identifying potential biological mechanisms that underlie social development, particularly when signals reach genome-wide significance. For example, we identified a variant within the *CADM2* gene as associated with LPB. Other variants in the *CADM2* gene have previously shown associations with NNCs, such as autism^{44,45}. These findings suggest that the biological role of genes at associated loci might be partially shared across the broader social behavioural continuum.

Even when aggregated across many variants, as reflected in the reported heritability estimates, the proportion of phenotypic variance in social behaviour attributable to genetic variation remains small. Following others (for example, ref. 27), we base our approach on the assumption that genetic factors represent only one of many contributions to complex social behaviour, alongside environmental, cultural, societal, regional, familial, peer group and individual influences. Thus, genetic effects are not deterministic, and social behaviour is shaped by many interacting factors.

Genetic effects may vary across different social situations, age or social characteristics (effect moderation), and genetic and non-genetic factors may become interlinked through gene–environment correlations. Consequently, at the aggregate level, genetic associations may not solely reflect biological mechanisms but also trait characteristics and their cultural context, and potential sources of bias. Thus, when studied with cohorts of predominantly European ancestry drawn from contemporary Western societies, the reported genetic associations reflect social behaviour as shaped by the developmental expectations of our Western society. Consequently, they may not generalize to non-European ancestries or cultural backgrounds, although some have cross-ancestry relevance, as shown here for LPB in African-ancestry populations.

PGS and genetic correlations, which capture the aggregate genetic effects of many small-effect variants, also proxy for relationships between traits. The genetic overlap will be stronger for social-behavioural outcomes that match for ancestry, social context and cultural backgrounds. However, as the overall genetic contributions are low, PGS are not suitable for predicting social outcomes for a single individual. As a result, individual-level PGS have little interpretative value.

This study provides both insight into the potential biological role of genes at associated loci through the identified GWAS signals and insight into the genetic factor structures and correlates of social behaviour, while accounting for social context, but without implying causality. Our findings can help better understand the developmental trajectories of social behaviour and may inform future longitudinal studies with prospective measurement collections.

The lack of genetic links between both the parent-reported (F1) and teacher-reported (F3) factors with the PSP factor (F4) is consistent with informant-specific characteristics, widely known in twin research¹⁶, including differences in knowledge about the young person, but also in social propensities elicited by different situations.

A key contribution of this study lies not only in the systematic modelling of heterogeneity in genetic effects underlying social behaviour but also, after extensive validation, in the systematic prediction of social context when studying genetic correlates. Genetic correlation analyses can provide insight into the studied trait through associated phenotypes (Box 1). We outline here how genetically predicted social context provides an interpretable framework for disentangling patterns of genetic overlap across complex traits, supporting the search for shared underlying developmental trajectories, including the timing and nature of processes relevant to mental health.

We identified strong PSP-specific links for a wide range of (near) adult externalizing, internalizing and lack-of-well-being measures, but also ADHD and MDD. These findings are consistent with complex behavioural networks and support research suggesting that children experiencing peer difficulties may be more aggressive, hyperactive and oppositional, but also more socially withdrawn and less sociable⁵⁶. Correlational patterns indicated the earlier emergence of links with PSP rather than LPB (if present), particularly for parent-reported GWAS. Consistently, genetic similarity across ADHD, MDD and PSP was identified with genetic CFA models.

In contrast, LPB was genetically related to a narrower range of adult population-based traits, despite greater power compared with PSP, based on h^2_{SNP} Z-scores. LPB captures only a single behavioural dimension, that is, reverse-coded prosocial abilities, and was primarily genetically associated with challenges in forming and maintaining

quality relationships, probably due to a reduced tendency to engage in helpful and sharing behaviours⁵⁷, but also genetically correlated with autism and schizophrenia. Genetic overlap across LPB and autism was detected using genetic CFA models, and links were weakly shared with schizophrenia. In particular, an earlier predicted onset of parent-reported LPB rather than PSP associations was observed for both later-life social outcomes and schizophrenia. In comparison, for autism, moderate-to-strong genetic correlations with both PSP and LPB, predicted to emerge from 4 years onwards across all reporters, matched the typical age of diagnosis⁵⁸, with social difficulties involving both a lack of sharing with others and problems in interacting with others⁵⁹. Thus, supported by similarities across multiple social correlates and the largely comparable power in social behaviour GWAS, the identified correlational patterns and structures are consistent with differences in developmental processes, where either PSP, LPB or both mark early stages in pathways leading to different later-life and mental health outcomes, without precluding bidirectional effects between PSP and LPB during development. However, neither LPB nor PSP were genetically linked to non-cognitive forms of EA⁶⁰, despite sufficient power, providing little evidence for a shared underlying genomic dimension.

This study demonstrates that social behaviour, as assessed with measures of LPB and PSP, is a diverse, multidimensional trait with quantitative changes in genetic effects (effect moderation) due to variation in social contextual factors, moving the field beyond qualitative reports^{9–11}. Our results highlight differences in genetic effects due to subphenotypic variation in social behaviour, contrasting person-driven (LPB) and interactive (PSP) trait characteristics (see earlier). Genetic heterogeneity due to differences in informants is consistent with differences in environment, capturing distinct social situations, and

possibly gene $\times$ environment interaction. Thus, genomic gene $\times$ environment approaches⁶¹ may have uncovered similar findings, although we observed only a few single-variant effects with non-overlapping 95% confidence intervals indicating reporter differences. Importantly, our findings highlight the presence of genetic effect moderation beyond environmental influences, especially developmental variation, which is known to shape genetic architectures⁶².

Adopting a scalable analytical framework for capturing the genomic basis of context-dependent traits, including developmental changes, allowed us to gain a deeper insight into the individual differences underlying social behaviour. More generally, we show that meta-regression approaches are capable of modelling and quantifying heterogeneity in genetic effects genome-wide, harmonizing cohort and phenotype information, especially through phenotypic mappings to the ICF²⁵, thus moving the field closer towards dynamic GWAS analyses. Deriving GWAS at different moderator levels, including age-related information, enabled the construction of meta-regression-derived PGS with context-specific accuracy, especially for European-ancestry cohorts. When discovery and target characteristics were matched, PGS captured up to a third of the h^2_{SNP} for social behaviour GWAS (incremental $R^2 \leq 1.32\%$), consistent with research showing that calibrating PGS analyses to different contexts enhances PGS accuracy²¹. Notably, social context could largely be inferred from the largest association signal, even with little information about target characteristics. Still, the prediction accuracy will depend on factors such as the studied population²¹, ancestry²¹ and also bias (for example, socially desirable responding³⁹) and was poor for self-reported LPB outcomes that were better predicted using parent-reported LPB PGS (Box 1).

Despite a considerable sample size capitalizing on repeat assessments, we detected only a few GWA loci, none of them matching previously identified candidate loci for social behaviour^{63–65}. The adjustment for repeated measures within cohorts (as part of sampling variance) might have been overly conservative, consistent with the lower power of meta-regression approaches to detect main effects in the absence of moderators²². Despite limitations, we demonstrated that meta-regression techniques are a powerful meta-analytic tool, enhancing both granularity and power of meta-GWAS beyond conventional meta-GWAS approaches. Expectedly, the predicted social behaviour GWAS will only partially proxy social-behavioural heterogeneity and additional moderators are likely to explain further variance, although an unspecific increase in the number of meta-regression moderators will lower statistical power. Although we find LPB cross-domain transferability to children of African ancestry, future studies will benefit from non-European-ancestry discovery populations.

In conclusion, we have shown that the genetic architecture of social behaviour is multidimensional and depends on social context, with underlying genetic effects reflecting variation in social domains, informants and age. Adopting a meta-regression approach, we demonstrated that heterogeneity in genetic association effects can be modelled and that genetic effect variation across social context is predictable, with implications for understanding the developmental pathways linking social behaviour to later-life outcomes and mental health.

Methods

Study cohorts

Within this study, we included data from 25 independent cohorts of European ancestry (73,321 individuals of European ancestry; Extended Data Table 1), embedded within the behaviour and cognition working group of the EAGLE Consortium²³ (<https://www.eagle-consortium.org/working-groups/behaviour-and-cognition/>). For each participating cohort, ethical approval was granted by their local research ethics committee, and written informed consent was given by the participants and/or their parents or legal guardians (Supplementary Note 1).

Phenotype definition

Prosocial behaviour measures were reverse-coded (LPB) to align with the PSP measures, where higher scores reflect more social behaviour difficulties. Where available, we studied repeatedly assessed (parent-, teacher- and self-reported) measures of PSP and LPB up to the age of 29 years ($N_{\text{obs}} = 491,246$; Supplementary Table 1).

Across the cohorts, dimensions of social behaviour (LPB and PSP) were predominantly assessed with the SDQ⁴ and the CBCL²⁴ (Supplementary Table 2). Both questionnaires are widely used, validated and standardized instruments assessing behavioural problems for research and clinical practice in both clinical and general population-based settings⁶⁶. All other instruments were mapped to either the LPB or PSP domain by aligning individual questionnaire items to the ICF²⁵ framework (Supplementary Note 2 and Supplementary Table 3).

Cohort-specific association analyses

Each cohort provided GWAS summary statistics and phenotype descriptive data on PSP and/or on prosocial behaviour (which was reverse-coded into LPB) assessed with parent-, teacher- and/or self-reports across ages 2–29 years (Supplementary Table 1).

Within each cohort, genotype and imputation quality control procedures were conducted, following standard procedures for GWAS (Supplementary Note 10 and Supplementary Table 36). Univariate GWAS analyses were performed using linear regression methods, assuming an additive linear model by regressing Z -standardized scores on posterior genotype probability while adjusting for covariate effects (sex, age, age² and ancestry-informative principal components). Further details on cohort-specific univariate GWAS are available in Supplementary Note 10.

Quality control of the GWAS summary statistics from all participating cohorts was carried out using the EasyQC⁶⁷ R package (R::EasyQC v.9.2), as described in detail in Supplementary Note 10. This included harmonization and filtering of SNPs, and alignment to the HRCr1.1 reference panel⁶⁸. The final number of retained SNPs for each cohort is available in Supplementary Table 36.

Genome-wide mixed-effects meta-regression analysis

For each SNP, genetic effects were meta-analysed across the 195 sets of GWAS summary statistics by implementing a mixed-effects meta-regression using the metafor R package (R::metafor v.3.0-2)²⁶. In contrast to a fixed-effects meta-analysis, a mixed-effects meta-regression model includes fixed effects (moderators) capturing systematic differences in SNP effects, here across domain (LPB or PSP), informant (parent-, teacher- or self-report) and age (2–29 years), thereby explicitly modelling and explaining structured heterogeneity in the genetic associations. In addition, residual heterogeneity not accounted for by the moderators is captured by random effects, here two random-effects terms, allowing for remaining between-cohort variability and between-GWAS variability (nested within cohorts). Sampling variability and dependence arising from repeated observations of overlapping individuals were not estimated directly but were approximated through the known sampling variance-covariance matrix V , which was supplied to the model (see later). Thus, this framework separates heterogeneity explained by moderators from unexplained residual heterogeneity, while appropriately accounting for non-independence among effect size estimates. The respective domain-, informant- and median age-related metadata, and phenotypic correlations across any repeat observations, were collected for each cohort (Supplementary Table 1).

For each SNP observed in cohort i and GWAS statistic j , variation (that is, heterogeneity) in genetic effects was modelled as a function of social domain, reporter, age and age² (moderators), while allowing for residual variation both between cohorts and between GWAS statistics nested within cohorts (equation (1)):

$$\hat{\beta}_{\text{SNP}_j} = \theta_0 + \theta_{\text{PSP}} x_{\text{PSP}_j} + \theta_{\text{Teacher}} x_{\text{Teacher}_j} + \theta_{\text{Self}} x_{\text{Self}_j} + \theta_{\text{Age}} x_{\text{Age}_j} + \theta_{\text{Age}^2} x_{\text{Age}_j}^2 + u_i + w_{ij} \quad (1)$$

where $\hat{\beta}_{\text{SNP}_j}$ is the estimated allelic effect for a given SNP for cohort i and GWAS statistic j . The intercept θ_0 represents the mean SNP effect for parent-reported LPB at the grand N -weighted mean age (10.14 years), with the age variable x_{Age_j} centred such that for each SNP the grand mean age corresponds to 0. The coefficient θ_{PSP} captures the difference in SNP effect for PSP, where x_{PSP} is a dummy variable coded 1 for PSP and 0 for LPB (reference category). Reporter effects were modelled with the coefficients θ_{Teacher} and θ_{Self} , representing the differences in SNP effects for teacher reports and self-reports, respectively, relative to parent reports (reference category). x_{Teacher} is a dummy-coded variable set to 1 for teacher reports and x_{Self} is a dummy-coded variable set to 1 for self-reports, and both are set to 0 for parent reports. The coefficients θ_{Age} and θ_{Age^2} model linear and quadratic age-related moderation, capturing differences in SNP effects for an increase of 1 year using the centred age variable x_{Age} . Random effects were specified using a nested structure, with each GWAS statistic j nested within cohort i . The cohort-specific random intercept u_i captures between-cohort heterogeneity and is distributed as $u_i \sim N(0, \tau_{\text{cohort}}^2)$. The within-cohort random intercept w_{ij} , capturing heterogeneity between GWAS statistics nested within the same cohort, is distributed as $w_{ij} \sim N(0, \tau_{\text{GWAS}}^2)$. The sum of τ_{cohort}^2 and τ_{GWAS}^2 represents the residual heterogeneity after accounting for moderators. SNP models were estimated using maximum likelihood.

The sampling variance-covariance matrix (V_i) within each cohort i was approximated, allowing for repeated observations for the same individuals (equation (2)). The adopted approach is analogous to methods accounting for correlated phylogenetic histories⁶⁹. Within each cohort i , S_i is a diagonal ($k \times k$) matrix containing the s.e. of the k effect sizes ($\hat{\beta}_{\text{SNP}}$) across k GWAS summary statistics, and Σ_{pheno_i} is the Pearson ($k \times k$) phenotypic correlation matrix across social behaviour measures within cohort i .

$$V_i = S_i \Sigma_{\text{pheno}_i}^{-1} S_i \quad (2)$$

The phenotypic correlation matrix (Σ_{pheno}) was adjusted for sample overlap following the formula by Bulik-Sullivan–Finucane and colleagues³⁵ (equation (3)), where $r_{m1,m2}$ is the phenotypic correlation between two measures m_1 and m_2 , $N_{m1,m2}$ is the number of overlapping samples measured in m_1 and m_2 , and N_{m1} and N_{m2} are the sample sizes of the respective measures.

$$\text{adj } r_{m1,m2} = r_{m1,m2} \times \frac{N_{m1,m2}}{\sqrt{N_{m1} \times N_{m2}}} \quad (3)$$

Assuming independence of sampling errors between cohorts, phenotypic correlations across measures from different cohorts were set to zero in the full phenotypic correlation matrix (Σ_{pheno}). Therefore, the full V matrix was constructed by combining V_i matrices, with all covariance terms across cohorts constrained to zero.

To ensure sufficient information for estimating the levels of the fixed-effect coefficients, analyses were restricted to SNPs that were available for both domains (PSP and LPB) for all 3 reporters (parent, teacher and self), with at least 3 different age time points and available in at least 2 cohorts, resulting in 7,852,971 SNPs.

To quantify the amount of heterogeneity explained by the moderators and to evaluate the model fit per SNP, we fitted a baseline that included only a fixed-effect intercept (θ_0) while retaining the same random-effects structure as specified in equation (1). We then compared the full meta-regression model (equation (1)) with this baseline (intercept-only) model using a likelihood-ratio test.

Deriving genome-wide genetic effects

For all SNPs for which the mixed-effects meta-regression model converged ($N_{\text{SNPs}} = 7,830,986$), we derived meta-analysed genetic effects across predefined moderator values, using the predict.rma function (R::metafor v.3.0-2)²⁶. For each SNP, we derived the meta-analysed genetic effects using the estimated coefficients of the corresponding SNP model ($\theta_0, \theta_{\text{PSP}}, \theta_{\text{Teacher}}, \theta_{\text{Self}}, \theta_{\text{Age}}, \theta_{\text{Age}^2}$). Specifically, we generated three groups of meta-analysed GWAS (Supplementary Table 4):

- (1) Low-resolution GWAS across social domains (two GWAS): A grand estimate was derived for each social domain ($\text{LPB}_{\text{All}}, \text{PSP}_{\text{All}}$). To do so, the social domain moderator value (x_{PSP}) was set to 0 for LPB and to 1 for PSP. x_{Age} and x_{Age^2} were set to the N -weighted mean across all 195 sets of GWAS (10.14 years old). Reporter moderator values ($x_{\text{Teacher}}, x_{\text{Self}}$) were set to the proportion of available data points for that SNP attributed to each reporter. For example, if a SNP has 80 total measures (50 parent, 20 teacher and 10 self), x_{Teacher} was set to 0.25 (20/80) and x_{Self} to 0.125 (10/80).
- (2) Medium-resolution GWAS across social domains and developmental stages (8 GWAS): GWAS were derived per social domain (LPB and PSP) and developmental window: early childhood (ages 2–5; $\text{LPB}_{\text{EarlyCh}}, \text{PSP}_{\text{EarlyCh}}$), mid-childhood (ages 6–11; $\text{LPB}_{\text{MidCh}}, \text{PSP}_{\text{MidCh}}$), adolescence (ages 12–17; $\text{LPB}_{\text{Adol}}, \text{PSP}_{\text{Adol}}$) and young adulthood (ages 18–29; $\text{LPB}_{\text{Adult}}, \text{PSP}_{\text{Adult}}$). For these GWAS, x_{Age} and x_{Age^2} were set to the N -weighted mean age for each developmental window across the 195 sets of GWAS. Specifically, this was 3.48 years for early childhood, 9.28 years for mid-childhood, 13.80 years for adolescence and 23.4 years for young adulthood. Reporter moderator values ($x_{\text{Teacher}}, x_{\text{Self}}$) were created as described for the low-resolution GWAS.
- (3) High-resolution GWAS across social domains, reporters and age (66 GWAS): GWAS were created based on a grid of moderator values that were contingent on the reported data across the 195 sets of GWAS (Fig. 1 and Supplementary Fig. 25). This grid included all possible combinations of social domain and reporter moderator values and discrete ages. To ensure reliability, we restricted the derivation of GWAS to age ranges with data from at least 2 cohorts or a single cohort with $N > 2,000$. Parent-reported LPB and PSP GWAS were derived from 2 to 17 years (16 ages $\times$ 2 domains = 32 GWAS). Teacher-reported LPB and PSP GWAS were derived from 7 to 13 years (7 ages $\times$ 2 domains = 14 GWAS). For self-reported LPB and PSP GWAS, the range was set from 9 to 18 years (10 ages $\times$ 2 domains = 20 GWAS).

Effective sample size

To estimate the effective sample size of derived GWAS summary statistics, we computed the effective sample size N_{eff} for each SNP following ref. 70 (equation (4)):

$$N_{\text{eff}} = \frac{(Z/\hat{\beta})^2}{2 \times \text{MAF} \times (1 - \text{MAF})} \quad (4)$$

where Z is the Z-statistic, $\hat{\beta}$ is the meta-analytic effect size at the selected moderator level and MAF is the minor allele frequency for that SNP.

Biological annotation of associated loci and mapped genes

Identification of GWA loci. SNP associations of the social behaviour GWAS were identified and annotated using Functional Mapping and Annotation of genetic associations (FUMA) software²⁸ (v.1.6.0). A genome-wide threshold of 5×10^{-8} was applied to identify lead SNPs. Note that no further correction for multiple testing was required, given that, for each SNP, GWAS summary statistics were derived from the

same meta-regression model. Independent GWAS signals were identified if they had a pairwise LD- $r^2 \leq 0.6$ using the 1000 Genomes Phase 3 European reference panel⁷¹. These SNPs were used in subsequent steps for gene mapping and gene prioritization. Lead SNPs were identified based on a pairwise LD- $r^2 \leq 0.1$ within an LD block distance of 250 kb. Furthermore, the major histocompatibility region was excluded from the analysis due to its extended LD structure.

Functional annotation and gene mapping. To map the independent SNPs to genes, both positional and eQTL mappings were used, as implemented in FUMA²⁸. Positional mapping was carried out using ANNOVAR annotations within FUMA and using a window of 10 kb. For eQTL mapping, SNPs were mapped to genes for SNPs that are likely to affect the expression of those genes up to 1 Mb, either in blood or brain tissues. Following default settings in FUMA, only SNP–gene pairs were considered, applying a false discovery rate of $<5\%$. Identified variants were annotated for their deleteriousness and mapped genes for their intolerance to functional mutations. Chromatin interaction mapping was carried out in FUMA; however, mapped genes were only used for gene-based analysis (GENE2FUNC in FUMA) and not for gene identification. Given that several associated GWA loci ($P < 5 \times 10^{-8}$) were observed outside the boundaries of gene coding sequences, and that a substantial part of common non-coding variant–trait association is mediated through regulatory effects⁷², xQTL mapping was conducted with MAGMA (v.1.10) software²⁹ to map variants within 35 kb upstream and 10 kb downstream per gene, including additional regulatory variant annotations (Supplementary Fig. 26). Quantitative trait loci affecting gene expression (eQTL), isoform proportion (isoQTL) and methylation (mQTL) patterns were annotated. In addition, we conducted a look-up of GWA loci for known genes and known phenotypic associations. All analyses are described in full in Supplementary Note 4.

Gene expression patterns, tissue specificity and molecular functions. Genes mapped by positional and eQTL mapping were annotated using GENE2FUNC (v.1.3.5) within FUMA for gene expression patterns and tissue specificity (differentially expressed genes) and overrepresentation in specific gene sets. In addition, we conducted a look-up of whole-brain cortical expression signatures of the genes with potential biological implications using the Allen Human Brain Atlas⁷³ (AHBA; <https://human.brain-map.org>; $N = 6$ post-mortem brains, aged 24–58 years). Moreover, we extracted gene expression data from the spatiotemporal gene expression dataset available from the BrainSpan Atlas of the Developing Human Brain⁷⁴ (<http://www.brainspan.org/>). These data cover transcriptome profiles of 16 different brain regions, spanning a time frame from embryonic development to late adulthood, for both males and females. All analyses are described in full in Supplementary Note 4.

Gene-based, gene-set and gene-property analyses. Gene-based, gene-set and gene-property analyses were conducted with MAGMA (v.1.08)²⁹. SNPs were mapped to protein-coding genes from Ensembl (build 85) using a window of 1 kb upstream and 0 kb downstream. A genome-wide, gene-based significance threshold was set to 2.62×10^{-6} ($0.05/19,039$ protein-coding genes tested). The multiple-testing adjusted P value threshold for gene-set analysis was set to 1.05×10^{-5} ($0.05/4,767$ gene sets) and the multiple-testing threshold for gene-property analyses was set to 4.03×10^{-4} ($0.05/124$ tissue types). A detailed description is provided in Supplementary Note 4.

GWAS selection for FUMA and MAGMA analysis. To provide a parsimonious overview of the biological insights from the 76 discovery GWAS summary statistics, FUMA and MAGMA analyses were carried out only for the strongest independent genome-wide hits (Supplementary Fig. 27). To identify the top hits, first, we retained all GWAS with at least 1 genome-wide signal ($N_{\text{GWAS}} = 22$). Second, across

these 22 GWAS, there were a total of 38 SNPs covering 6 independent genomic loci (Table 1). For each of the 6 loci, we interrogated the GWAS with the strongest association signal; specifically, these were self-reported LPB at 9 years, parent-reported LPB at 4 and 6 years, parent-reported PSP at 2 and 6 years and early childhood PSP (Supplementary Fig. 27).

Conditional analyses. To confirm the independence of genome-wide hits, we conducted conditional analysis using GCTA-COJO²⁷ (GCTA v.1.94) on the 6 selected GWAS, where we conditioned the association of the lead SNP on the other SNPs that passed the genome-wide threshold ($P < 5 \times 10^{-8}$) and were located on the same chromosome.

Genetic architecture of social behaviour

SNP heritability and genetic correlation analysis. To study the overarching genetic architecture of the derived GWAS summary statistics (Supplementary Table 4), we estimated h^2_{SNP} and r_g using LDSC regression (v.3)³⁴ and LDSC correlation³⁵, as described in Supplementary Note 5. All analyses were based on HapMap 3 SNPs only, and precalculated LD scores from the European 1000 Genomes reference cohort⁷¹ were used.

Genomic factor analysis across social behaviour. To study the relationships across the 66 discovery GWAS summary statistics (high-resolution: across social domains, reporters and age), we carried out a data-driven and hypothesis-free approach, following a similar pipeline as applied in our previous work³⁶. The aim of this approach is to identify clusters of variables sharing genetic variation. Specifically, we carried out a combination³⁶ of principal component analysis, exploratory factor analysis and CFA, where the CFA model is fitted with GenomicSEM³⁷ software, an approach that has been described in detail in Supplementary Note 5.

Polygenic prediction of social behavioural differences

PGS for LPB and PSP across reporters and ages were constructed in 4 independent cohorts ($N_{\text{ind}} = 16,305$), described in the Supplementary Notes (Supplementary Notes 6 and 7), including extensive validation analyses in the MCS (ages 3–17, 18 phenotypes)³⁸ and further analyses in 2 European-ancestry cohorts and 1 African-ancestry cohort (Supplementary Note 6 and Supplementary Table 15). Detailed information on phenotype and genetic data, consent and ethical approval for each of these cohorts is available in Supplementary Note 6 and Supplementary Table 15.

PGS were calculated with PRS-CS⁷⁵ (version June 2021) for European-ancestry cohorts and PRS-CSx software⁷⁶ (version November 2024) for the admixed African-ancestry cohort. Using the re-estimated effect sizes, PGS were generated with PLINK⁷⁷ (v.1.9) and subsequently Z-standardized. PGS were used as predictors in linear models of quantitative outcomes, which were adjusted for covariates (age, sex and the first ten ancestry-informative principal components). The variation explained by the PGS was captured as incremental R^2 . All analyses were corrected for multiple testing, as detailed in Supplementary Note 7.

Genetic correlation analyses with population-based traits and NNCs

To study the genetic overlap of social behaviour with other population-based traits ($N = 51$) and clinically defined NNCs ($N = 5$, ADHD, autism, MDD, schizophrenia and BD), we computed bivariate genetic correlations using LDSC software³⁵. Population-based traits included a broad range of developmental and adult phenotypes, including externalizing-related, internalizing-related, social, well-being, personality, cognition and EA, language, and anthropometric and birth outcomes (Supplementary Table 22). Respective summary statistics were obtained from existing publications (Supplementary Table 22) or the UK Biobank GWAS repository (<http://www.nealelab.is/uk-biobank/>). The multiple-testing threshold (Bonferroni adjusted) was set to

$P < 8.89 \times 10^{-4}$ (0.05/56 traits). Within this study, we refer to the strength of a correlation when reporting the P value and refer to the magnitude of the correlation when reporting the size of the correlation, which can be low (<0.3), moderate (0.3–0.6) or strong (>0.6).

Genomic factor analysis across social behaviour and NNCs

In addition to genetic correlations, we studied the multivariate architecture across the five NNCs and social behaviour. Specifically, we used the meta-regression-derived medium-resolution social behaviour GWAS differentiating across social domains and developmental stages with sufficient power (h^2_{SNP} Z-score > 3 , $\text{LPB}_{\text{EarlyCh}}$, $\text{PSP}_{\text{EarlyCh}}$, $\text{LPB}_{\text{MidCh}}$, $\text{PSP}_{\text{MidCh}}$, LPB_{Adol} , PSP_{Adol}) and clinically defined NNCs as studied for correlation analyses (Supplementary Table 22): ADHD, autism, MDD, schizophrenia and BD (Supplementary Note 8). We refrained from including further NNCs into the model to ensure computational feasibility of fitting a model across traits with differing h^2_{SNP} estimates (h^2_{SNP} range, 2.3–21%). To identify a genomic factor model, we adopted a data-driven approach³⁶ using a combination of principal component analysis, exploratory factor analysis and CFA, where the CFA model was fitted with GenomicSEM³⁷ software, as described earlier (Supplementary Note 5). In addition, to further characterize the identified NNC factors, we subsequently mapped EA onto the factor structure of the best-fitting CFA model Supplementary Note 8.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Social behaviour GWAS: meta-regression-derived GWAS summary statistics are available via the MPI Archive (<https://hdl.handle.net/1839/7a2d6cfd-431a-4e88-82f5-82ee57060958>). 1958BC: phenotype and genetic data are provided by the UK Data Service at the University of Essex after registration (<https://doi.org/10.5255/UKDA-Series-2000032>). Information about data access is described in the CLS Data Access Framework at https://cls.ucl.ac.uk/wp-content/uploads/2017/02/CLS_Data_Access_Framework.pdf. ABCD-NL: due to ethical considerations, the individual data from the ABCD study cannot be deposited in a public repository. Any researcher can request the data by submitting a general publication proposal form to the ABCD Project Team as outlined at <https://www.amc.nl/web/abcd-studie-2/abcd-studie/voor-onderzoekers.htm> under the section ‘use of data’ (Dutch: *gebruik van data*), by email: abcd@amsterdamumc.nl. Our data access policy can be found at <https://www.amc.nl/web/abcd-studie-2/abcd-studie/voor-onderzoekers.htm>. ABCD-US: the ABCD data used in this study came from the ABCD Data Release 4.0. The ABCD direct data access via DNA was discontinued in 2025 and the ABCD archive moved to the NIH Brain Development Cohorts (HBDC) Data Hub. Access to ABCD data can be obtained via the NBDC (<https://www.nbdc-datahub.org/>). ALSPAC: the data used are available through a fully searchable data dictionary (<http://www.bristol.ac.uk/alspac/researchers/our-data/>). Access to ALSPAC data can be obtained as described within the ALSPAC data access policy (<http://www.bristol.ac.uk/alspac/researchers/access/>). BREATHE: access to BREATHE data can be obtained by contacting jordi.sunyer@isglobal.org or mariona.bustamante@isglobal.org. CATSS: the data used from the CATSS cohort are not publicly available to protect sensitive phenotype information for participating children. Access to CATSS can be obtained by submitting a data request. COPSAC: data from the COPSAC cohort is classified as sensitive, personally identifiable data and thus cannot be made publicly available, but can potentially be made available upon request under a data transfer agreement. DCHS: the DCHS is committed to the principle of data sharing. De-identified data will be made available to requesting researchers as appropriate. Requests for collaborations are welcome. More information can be found on our website at

<http://www.paediatrics.uct.ac.za/scab/dclhs>. ELVS: the data used from the ELVS dataset are not readily available due to the ethical governance set in place for this study. The data are shared following application to investigators of the ELVS study. Requests to access the genetic datasets should be directed to Angela Morgan, angela.morgan@mcri.edu.au. FinnTwin12: in accordance with the Finnish Biobank Act, the data used in the analysis are deposited in the Biobank of the Finnish Institute for Health and Welfare (<https://thl.fi/en/our-services/thl-biobank1>). They are available to researchers from academia and companies after a written application and following the relevant Finnish legislation. To ensure the protection of privacy and compliance with national data protection legislation, a data use/transfer agreement is needed, the content and specific clauses of which will depend on the nature of the requested data. GenR: data from the Generation R cohort are not publicly available due to legal and ethical restrictions; however, access can be requested via the Generation R administration (secretariaat.genr@erasmusmc.nl). Requests for Generation R data access are evaluated by the Generation R Management Team. GINILISA: due to data protection reasons, the datasets generated and/or analysed during the current study cannot be made publicly available. The datasets are available to interested researchers from the corresponding author upon reasonable request (for example, reproducibility), provided the release is consistent with the consent given by the GINplus and LISA study participants. Ethical approval might be obtained for the release, and a data transfer agreement from the legal department of the Helmholtz Zentrum München must be accepted. GLAKU: researchers can obtain a de-identified GLAKU dataset after having obtained approval from the GLAKU Study Board. HRS: the phenotype survey data are provided by the HRS without charge, after registration. The genomic data are provided by the health and retirement study after institutional review board approval by NCBI-dbGaP (dbGaP study accession: [pbs000428.v2.p2](https://www.ncbi.nlm.nih.gov/studies/pbs000428.v2.p2)), without charge. Sensitive data, including genetic information and biomarkers, are only available through an approved data application process (Health and Retirement Study, DUA Review Committee, 426 Thompson Street, Ann Arbor, Michigan 48104-2321, hdsdatareq@umich.edu). By requesting access to the HRS genetic information via NCBI-dbGaP, the authors agreed that the data are only shared within the working group permitted to work with the data (St Pourcain group; <https://hrs.isr.umich.edu/sites/default/files/HRS-Genetic-Data-Access-Agreement.pdf>). Information about the data access is described at: <https://hrs.isr.umich.edu/data-products/>. INMA: INMA data (INMA-GSA and INMA-Omni1) are available by request from the Infancia y Medio Ambiente Executive Committee for researchers who meet the criteria for access to confidential data. Access to INMA data can be obtained by contacting inma@proyectoinma.org. INSchool: Raw data from INSchool are not publicly available because of limitations in ethical approvals, and the summary data will be available by contacting marta.ribases@vhir.org. LSAC: data from the LSAC cohort can be obtained by completing an online form, following instructions from the LSAC data custodian, The Australian Institute of Family Studies. Information about data access is described at: <https://growingupinaustralia.gov.au/data-and-documentation/accessing-lsac-data>. MCS: phenotype and genetic data from the MCS are provided by the UK Data Service at the University of Essex after registration (<https://doi.org/10.5255/UKDA-Series-2000031>). Information about data access is described in the CLS Data Access Framework: https://cls.ucl.ac.uk/wp-content/uploads/2017/02/CLS_Data_Access_Framework.pdf. MoBa: data from the Norwegian Mother, Father and Child Cohort Study and the Medical Birth Registry of Norway used in this study are managed by the national health register holders in Norway (Norwegian Institute of Public Health). The consent given by the participants does not allow for storage of data on an individual level in repositories or journals. Researchers who want access to datasets for replication should apply through helsedata.no. Access to datasets requires approval from the Regional Committee for Medical and Health Research Ethics in Norway (REC)

and an agreement with MoBa. In addition, compliance with the EU General Data Protection Regulation is required. MUSP: MUSP data are available from a third party upon reasonable request. Contact details can be found at: <https://social-science.uq.edu.au/mater-university-queensland-study-pregnancy?p=5#5>. NFBC1986: NFBC data are available from the University of Oulu, Infrastructure for Population Studies. Permission to use the data can be applied for research purposes via an electronic material request portal. In the use of data, we follow the EU General Data Protection Regulation (679/2016) and the Finnish Data Protection Act. The use of personal data is based on a cohort participant's written informed consent in their latest follow-up study, which may cause limitations to its use. Contact the NFBC project centre (NFBCprojectcenter@oulu.fi) and visit the cohort website (www.oulu.fi/nfbc) for more information. NTR: the data used from the NTR cohort are not publicly available to protect sensitive phenotype information for participating children. Access to NTR can be obtained by submitting a data request. The Raine Study: restrictions apply to the availability of data generated or analysed during this study to preserve patient confidentiality. On request (<https://ross.rainestudy.org.au>) details of the restrictions and any conditions under which data access is granted may be provided. SYS: SYS data are available upon request and should be addressed to Dr Zdenka Pausova (zdenka.pausova@sickkids.ca) and Dr Tomas Paus (tpausresearch@gmail.com). Further details about the protocol can be found at <https://saguenay-youth-study.net/>. TEDS: researchers can apply for access to the TEDS data through their data access mechanism (www.teds.ac.uk/researchers/teds-data-access-policy). TRAILS: TRAILS data are not publicly available to researchers outside the TRAILS consortium, but can be obtained by submitting a publication proposal. For more details, please see www.trails.nl.

Code availability

This study used openly available software and codes, specifically abagen toolbox (v.0.1.1, <https://github.com/rmarkello/abagen>), FUMA (v.1.6.0, <https://fuma.ctglab.nl/>), GCTA-COJO (v.1.94, <https://yanglab.westlake.edu.cn/software/gcta/>), GEMMA (<https://github.com/genetics-statistics/GEMMA>), multivariate GWAMA (https://github.com/baselmans/multivariate_GWAMA), LDSC (v.1.0.1, <https://github.com/bulik/ldsc>), METAL (v.2011-03-25, <https://csg.sph.umich.edu/abecasis/metal/>), MAGMA (v.1.08 and v.1.10, <https://ctg.cncr.nl/software/magma>), PLINK (v.1.9, <https://www.cog-genomics.org/plink/1.9/>), ProbABEL (<https://github.com/GenABEL-Project/ProbABEL>), PRSs (4 June 2021, <https://github.com/getian107/PRSs>), PRSs (v.1.0, <https://github.com/getian107/PRSs>) and SNPTEST (https://mathgen.stats.ox.ac.uk/genetics_software/snpTEST/snpTEST_v2.4.1.html). We used the following R packages: cerebioviz v.1.0, EasyQC v.9.2, fsbrain v.0.5.5, GenomicSEM v.0.05, lavaan v.0.6-10, LDlinkR v.1.4.0, LocusZoom v.2.1, Matrix v.1.7-0, metafor v.3.0-2 and nFactors v.2.4.1. We accessed the following publicly available databases: GWAS Catalog (<https://www.ebi.ac.uk/gwas/>), pheWeb (<https://pheweb.org/>) and SFARI (<https://gene.sfari.org/database/human-gene/>). Scripts used in this study are available at <https://gitlab.gwdg.de/pghc/social-behaviour/social-behaviour-gwas-2026>.

References

- Junge, C., Valkenburg, P. M., Deković, M. & Branje, S. The building blocks of social competence: contributions of the Consortium of Individual Development. *Dev. Cogn. Neurosci.* **45**, 100861 (2020).
- Mace, R. in *Social Behaviour: Genes, Ecology and Evolution* (eds Moore, A. J. et al.) 395–409 (Cambridge Univ. Press, 2010); <https://doi.org/10.1017/CBO9780511781360.031>
- Brownell, C. A. & The Early Social Development Research Lab. Prosocial behavior in infancy: the role of socialization. *Child Dev. Perspect.* **10**, 222–227 (2016).
- Goodman, R. The Strengths and Difficulties Questionnaire: a research note. *J. Child Psychol. Psychiatry* **38**, 581–586 (1997).
- Fabiano, G. A., Vujnovic, R. K. & Pariseau, M. E. in *Handbook of Clinical Psychology Competencies* (eds Thomas, J. C. & Hersen, M.) 1563–1588 (Springer, 2010); https://doi.org/10.1007/978-0-387-09757-2_57
- Hartup, W. W. & Stevens, N. Friendships and adaptation across the life span. *Curr. Dir. Psychol. Sci.* **8**, 76–79 (1999).
- Agostin, T. M. & Bain, S. K. Predicting early school success with developmental and social skills screeners. *Psychol. Sch.* **34**, 219–228 (1997).
- Reef, J., van Meurs, I., Verhulst, F. C. & van der Ende, J. Children's problems predict adults' DSM-IV disorders across 24 years. *J. Am. Acad. Child Adolesc. Psychiatry* **49**, 1117–1124 (2010).
- Knafo-Noam, A., Vertsberger, D. & Israel, S. Genetic and environmental contributions to children's prosocial behavior: brief review and new evidence from a reanalysis of experimental twin data. *Curr. Opin. Psychol.* **20**, 60–65 (2018).
- Knafo, A. & Plomin, R. Prosocial behavior from early to middle childhood: genetic and environmental influences on stability and change. *Dev. Psychol.* **42**, 771–786 (2006).
- St Pourcain, B. et al. Heritability and genome-wide analyses of problematic peer relationships during childhood and adolescence. *Hum. Genet.* **134**, 539–551 (2015).
- Lewis, G. J., Haworth, C. M. & Plomin, R. Identical genetic influences underpin behavior problems in adolescence and basic traits of personality. *J. Child Psychol. Psychiatry* **55**, 865–875 (2014).
- Fedko, I. O. et al. Heritability of behavioral problems in 7-year olds based on shared and unique aspects of parental views. *Behav. Genet.* **47**, 152–163 (2017).
- Pappa, I. et al. Single nucleotide polymorphism heritability of behavior problems in childhood: genome-wide complex trait analysis. *J. Am. Acad. Child Adolesc. Psychiatry* **54**, 737–744 (2015).
- Knafo-Noam, A., Uzefovsky, F., Israel, S., Davidov, M. & Zahn-Waxler, C. The prosocial personality and its facets: genetic and environmental architecture of mother-reported behavior of 7-year-old twins. *Front. Psychol.* **6**, 112 (2015).
- Kan, K.-J., van Beijsterveldt, C. E. M., Bartels, M. & Boomsma, D. I. Assessing genetic influences on behavior: informant and context dependency as illustrated by the analysis of attention problems. *Behav. Genet.* **44**, 326–336 (2014).
- Bralten, J. et al. Genetic underpinnings of sociability in the general population. *Neuropsychopharmacology* **46**, 1627–1634 (2021).
- Warrier, V., 23andMe Research Team, Bourgeron, T. & Baron-Cohen, S. Genome-wide association study of social relationship satisfaction: significant loci and correlations with psychiatric conditions. Preprint at *bioRxiv* <https://doi.org/10.1101/196071> (2017).
- Day, F. R., Ong, K. K. & Perry, J. R. B. Elucidating the genetic basis of social interaction and isolation. *Nat. Commun.* **9**, 2457 (2018).
- Socrates, A. J. et al. Polygenic risk of social isolation behavior and its influence on psychopathology and personality. *Mol. Psychiatry* **29**, 3599–3606 (2024).
- Hou, K. et al. Calibrated prediction intervals for polygenic scores across diverse contexts. *Nat. Genet.* **56**, 1386–1396 (2024).
- Pagoni, P. et al. Meta-regression of genome-wide association studies to estimate age-varying genetic effects. *Eur. J. Epidemiol.* **39**, 257–270 (2024).
- Middeldorp, C. M. et al. The Early Growth Genetics (EGG) and EARly Genetics and Lifecourse Epidemiology (EAGLE) consortia: design, results and future prospects. *Eur. J. Epidemiol.* **34**, 279–300 (2019).
- Achenbach, T. M. *Manual for the Child Behaviour Check-List/4-18 and 1991 Profile* (Univ. of Vermont, 1991).

25. World Health Organization. *International Classification of Functioning, Disability and Health: Children and Youth Version: ICF-CY* (World Health Organization, 2007).
26. Viechtbauer, W. Conducting meta-analyses in R with the metafor package. *J. Stat. Softw.* **36**, 1–48 (2010).
27. Yang, J. et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375 (2012).
28. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
29. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).
30. Fogel, A. I. et al. SynCAMs organize synapses through heterophilic adhesion. *J. Neurosci.* **27**, 12516–12530 (2007).
31. Rudebeck, P. H. & Rich, E. L. Primer: the orbitofrontal cortex. *Curr. Biol.* **28**, R1083–R1088 (2018).
32. Ng, B. et al. An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. *Nat. Neurosci.* **20**, 1418–1426 (2017).
33. Gandal, M. J. et al. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* **362**, eaat8127 (2018).
34. Bulik-Sullivan, B. K. et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
35. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
36. de Hoyos, L. et al. Structural models of genome-wide covariance identify multiple common dimensions in autism. *Nat. Commun.* **15**, 1770 (2024).
37. Grotzinger, A. D. et al. Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nat. Hum. Behav.* **3**, 513–525 (2019).
38. Connelly, R. & Platt, L. Cohort profile: UK Millennium Cohort Study (MCS). *Int. J. Epidemiol.* **43**, 1719–1725 (2014).
39. Lee, Z. & Sargeant, A. Dealing with social desirability bias: an application to charitable giving. *Eur. J. Mark.* **45**, 703–719 (2011).
40. Wing, L. & Gould, J. Severe impairments of social interaction and associated abnormalities in children: epidemiology and classification. *J. Autism Dev. Disord.* **9**, 11–29 (1979).
41. Schlag, F. et al. Polygenic risk for mental disorder reveals distinct association profiles across social behaviour in the general population. *Mol. Psychiatry* **27**, 1588–1598 (2022).
42. Baselmans, B. M. L. et al. Multivariate genome-wide analyses of the well-being spectrum. *Nat. Genet.* **51**, 445–451 (2019).
43. Pasman, J. A. et al. The CADM2 gene and behavior: a phenome-wide scan in UK-Biobank. *Behav. Genet.* **52**, 306–314 (2022).
44. Casey, J. P. et al. A novel approach of homozygous haplotype sharing identifies candidate genes in autism spectrum disorder. *Hum. Genet.* **131**, 565–579 (2012).
45. Feliciano, P. et al. Exome sequencing of 457 autism families recruited online provides evidence for autism risk genes. *npj Genom. Med.* **4**, 19 (2019).
46. Strawbridge, R. J. et al. Genetics of self-reported risk-taking behaviour, trans-ethnic consistency and relevance to brain gene expression. *Transl. Psychiatry* **8**, 178 (2018).
47. Liu, M. et al. 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 (2019).
48. Karlsson Linnér, R. et al. Multivariate analysis of 1.5 million people identifies genetic associations with traits related to self-regulation and addiction. *Nat. Neurosci.* **24**, 1367–1376 (2021).
49. Okbay, A. et al. Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. *Nat. Genet.* **54**, 437–449 (2022).
50. Davies, G. et al. Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nat. Commun.* **9**, 2098 (2018).
51. Martinelli, S. et al. Heterozygous germline mutations in the CBL tumor-suppressor gene cause a Noonan syndrome-like phenotype. *Am. J. Hum. Genet.* **87**, 250–257 (2010).
52. Balak, C. et al. Rare de novo missense variants in RNA helicase DDX6 cause intellectual disability and dysmorphic features and lead to P-body defects and RNA dysregulation. *Am. J. Hum. Genet.* **105**, 509–525 (2019).
53. Ferguson, B. R. & Gao, W.-J. Thalamic control of cognition and social behavior via regulation of GABAergic signaling and E/I balance in the medial prefrontal cortex. *Biol. Psychiatry* **83**, 657–669 (2018).
54. Ritter, A. L. et al. Expanding the phenotypic spectrum of ARCN1-related syndrome. *Genet. Med.* **24**, 1227–1237 (2022).
55. Warrier, V. et al. Genome-wide meta-analysis of cognitive empathy: heritability, and correlates with sex, neuropsychiatric conditions and cognition. *Mol. Psychiatry* **23**, 1402–1409 (2018).
56. Boivin, M. The origin of peer relationship difficulties in early childhood and their impact on children's psychosocial adjustment and development. In *Encyclopedia of Early Child Development* (eds Tremblay R. E. et al.) 1–7 (Centre of Excellence for Early Childhood Development, 2005); <http://www.child-encyclopedia.com/documents/BoivinANGxp.pdf>
57. Brownell, C. A. Early development of prosocial behavior: current perspectives. *Infancy* **18**, 1–9 (2013).
58. van't Hof, M. et al. Age at autism spectrum disorder diagnosis: a systematic review and meta-analysis from 2012 to 2019. *Autism* **25**, 862–873 (2021).
59. Lord, C. et al. Autism spectrum disorder. *Nat. Rev. Dis. Prim.* **6**, 5 (2020).
60. Demange, P. A. et al. Investigating the genetic architecture of noncognitive skills using GWAS-by-subtraction. *Nat. Genet.* **53**, 35–44 (2021).
61. Young, A. I., Wauthier, F. L. & Donnelly, P. Identifying loci affecting trait variability and detecting interactions in genome-wide association studies. *Nat. Genet.* **50**, 1608–1614 (2018).
62. Visscher, P. M., Hill, W. G. & Wray, N. R. Heritability in the genomics era—concepts and misconceptions. *Nat. Rev. Genet.* **9**, 255–266 (2008).
63. Tanksley, P. T., Motz, R. T., Kail, R. M., Barnes, J. C. & Liu, H. The genome-wide study of human social behavior and its application in sociology. *Front. Sociol.* **4**, 53 (2019).
64. Chabris, C. F. et al. Why it is hard to find genes associated with social science traits: theoretical and empirical considerations. *Am. J. Public Health* **103**, S152–S166 (2013).
65. Holmes, A. The neuroscience and genomics of social behavior. *Genes Brain Behav.* **18**, e12551 (2019).
66. Warnick, E. M., Bracken, M. B. & Kasl, S. Screening efficiency of the Child Behavior Checklist and Strengths and Difficulties Questionnaire: a systematic review. *Child Adolesc. Ment. Health* **13**, 140–147 (2008).
67. Winkler, T. W. et al. Quality control and conduct of genome-wide association meta-analyses. *Nat. Protoc.* **9**, 1192–1212 (2014).
68. McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
69. Lajeunesse, M. J. Meta-analysis and the comparative phylogenetic method. *Am. Naturalist* **174**, 369–381 (2009).
70. Mallard, T. T. et al. Multivariate GWAS of psychiatric disorders and their cardinal symptoms reveal two dimensions of cross-cutting genetic liabilities. *Cell Genomics* **2**, 100140 (2022).

71. 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
72. Maurano, M. T. et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).
73. Gryglewski, G. et al. Spatial analysis and high resolution mapping of the human whole-brain transcriptome for integrative analysis in neuroimaging. *NeuroImage* **176**, 259–267 (2018).
74. BrainSpan Consortium. BrainSpan: Atlas of the Developing Human Brain. [Dataset] brainspan.org (2011).
75. Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C. A. & Smoller, J. W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.* **10**, 1776 (2019).
76. Ruan, Y. et al. Improving polygenic prediction in ancestrally diverse populations. *Nat. Genet.* **54**, 573–580 (2022).
77. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, 7 (2015).

Acknowledgements

We thank all the children, twins, families and participants who took part in, or are taking part in, the 29 cohorts whose data contributed to the GWAS meta-analyses or follow-up analysis for making this work possible. We also thank all employees of the cohorts included in this study. This project was embedded within the EAGLE Consortium, and follow-up analyses were also embedded in the R2D2 Consortium. Cohort-specific acknowledgements and funding information can be found in Supplementary Note 11. In addition, we thank all cohorts and researchers who made their summary statistics available to us. This includes the GenLang Consortium, Social Science Genetic Association Consortium, Early Growth Genetics Consortium, EAGLE Consortium and the Psychiatric Genomics Consortium. We extend our thanks and acknowledge the input of the R2D2-MH Adult Co-Creation Group for their support and contribution to the design of the study materials. We thank W. Viechtbauer for helpful discussions about conducting meta-regressions using metafor software.

Author contributions

B.S.P. conceptualized and designed the meta-regression study. B.S.P. and L.d.H. implemented the meta-regression design. L.d.H. performed the meta-regression analysis and generated the predicted social behaviour GWAS summary statistics. B.S.P. supervised the research. F.S., L.d.H., D.A.G.A. and N.Y.T.N. were involved in data acquisition and collation of cohort-level summary statistics, and in processing the quality control of these. The following authors performed the data analysis within cohorts contributing to the meta-analysis: E.M. (ABCD-NL), F.S. (ABCD-US and ALSPAC), L.d.H. (ABCD-US and ALSPAC), D.A.G.A. (ALSPAC), N.V.-T. (BREATHE), E.D.R. (CATSS), C.-E.T.P. (COPSAC), E.E. (ELVS), T. Palviainen (FinnTwin12), A.N. (GenR), E.T. (GINILISA), P.G. (GLAKU), N.F.-J. (INMA_GSA), S.A. (INMA_Omni1), N.L. (INSchool), K.L. (LSAC), E.C.C. (MoBa), P.C. (NFBC1986), E.L.d.Z. (NTR), C.A.W. (Raine), Z.L. (SYS), A.G.A. (TEDS) and I.M.N. (TRAILS). The following authors are the principal investigators of the cohorts contributing to the meta-analysis: T.G.M.V. (ABCD-NL), B.S.P. (ABCD-US and ALSPAC), J. Sunyer (BREATHE), H.L. (CATSS), K.B. (COPSAC), A.T.M. (ELVS), J.K. (FinnTwin12), C.A.M.C. (GenR), M.S. (GINILISA), J.L. (GLAKU), J.R.B. (INMA_GSA), M.B. (INMA_Omni1), J.A.R.-Q. (INSchool), M.R. (INSchool), M.C. (INSchool), R.B. (INSchool), C.M.M. (MUSP), A. Havdahl (MoBa), S. Sebert (NFBC1986), D.I.B. (NTR), A.J.O.W. (Raine), T. Paus (SYS), K. Rimfeld (TEDS) and C.A.H. (TRAILS). The following authors supported the analysis or interpretation of data of the cohorts contributing to the meta-analysis: E. Verhoef (ALSPAC), A.E.T. (CATSS), P.L. (CATSS), M.H.-L. (COPSAC), T.S.A. (COPSAC), S.R. (ELVS), E. Vuoksimaa (FinnTwin12), H.T. (GenR), M.L. (GenR), V.W.J. (GenR), A.M. (GLAKU), K. Räikkönen (GLAKU), S.M. (INMA_GSA), S.L. (INMA_Omni1), M.S.A. (INSchool), M.W. (LSAC), E.Y. (MoBa), T.R.-K.

(MoBa), A. Rodriguez (NFBC1986), A. Heiskala (NFBC1986), M.-R.J. (NFBC1986), C.E.P. (Raine), J. Shin (SYS), Z.P. (SYS), A. Ronald (TEDS), R.P. (TEDS) and A.J.O. (TRAILS). D.A.G.A., M.H.B., B.S.P. and S.B. performed the ICF analysis. H.S., A.K. and A.J. performed the brain expression look-up analysis in Allen Human Brain Atlas and BrainSpan, and C.E. supervised the research. L.d.H. performed the biological annotation analysis in FUMA, the look-up analysis in GWAScatalog and pheWeb, and the genetic architecture analysis. E.B. and L.C.M. performed the biological annotation analysis in MAGMA, and T.R. and T.B. supervised the research. L.d.H. performed the PGS analysis in the MCS, 1958 birth cohort (1958BC) and Health and Retirement Study (HRS) cohorts. D.A.G.A. prepared the phenotypic data for the 1958BC. S. Scatolin prepared the genetic and phenotypic data for the HRS. S.v.d.B. prepared the genetic and phenotypic data for the MCS. M.L.C. performed the PGS analysis in the Drakenstein Child Health Study (DCHS) cohort, and M.T.L., S.-M.K. and D.J.S. supervised the research. F.S., L.d.H., S.J. and B.S.P. curated the preselected population-based traits and neurodevelopmental conditions to run genetic correlations. L.d.H. performed the genetic correlation analysis. S.J. performed the GenomicSEM analysis with neurodevelopmental conditions. F.S. designed the pipeline to run conventional meta-analysis in METAL and GWAMA, and F.S. and L.d.H. performed the analysis. L.d.H. and B.S.P. drafted the paper. J.K.B., S.E.F. and V.W. critically reviewed the paper. All authors contributed to writing and provided feedback on the paper.

Funding

L.d.H., F.S., D.A.G.A., S.J., E. Verhoef, S.v.d.B., S. Scatolin, E.E., N.Y.T.N., S.E.F. and B.S.P. were supported by the Max Planck Society. S.J., A.K., A.J., E.B., L.C.M., H.S., M.H.B., T.R., C.E., S.B., T.B. and B.S.P. were supported by the R2D2-MH, which has been funded by Horizon Europe (grant agreement number 101057385), by UK Research and Innovation (UKRI) under the UK government's Horizon Europe funding guarantee (grant number 10039383) and by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number 22.00277. B.S.P. is also supported by the Radboud University Donders RSF 2025. T.G.M.V. was supported by ZonMW (TOP 40-00812-98-11010). C.A.M.C. is supported by the Horizon Europe Research and Innovation Programme (FAMILY, grant agreement number 101057529, also to A.N.; HappyMums, grant agreement number 101057390) and the European Research Council (TEMPO; grant agreement number 101039672, also to A.N.). A.N. and H.T. are supported by a grant from the Dutch Ministry of Education, Culture, and Science and the Netherlands Organization for Scientific Research (NWO grant number 024.001.003, Consortium on Individual Development). The work of H.T. is further supported by the European Union's Horizon 2020 research and innovation programme (contract grant number 633595, DynaHealth) and an NWO-VICI grant (NWO-ZonMW; 016.VICI.170.200). M.L. was supported by a scholarship from the China Scholarship Council (201706990036). This work of M.L. was further supported by the Ter Meulen Grant of the Royal Netherlands Academy of Arts and Sciences (KNAW). J.L. has received research support from the Strategic Research Council established within the Academy of Finland (decision number 352700). S.A. acknowledges her Miguel Servet contract (CP22/00026) awarded by the Instituto de Salud Carlos III and co-funded by the European Union fund (Fondo Social Europeo Plus, FSE+). A. Havdahl was supported by funding from the Research Council of Norway (RCN) (numbers 274611 and 336085) and the South-Eastern Norway Regional Health Authority (HSØ; numbers 2020022 and 2018059). E.C.C. was supported by funding from the RCN (number 274611) and HSØ (number 2021045) and is a member of the MRC Integrative Epidemiology Unit at the University of Bristol, which is supported by the Medical Research Council and the University of Bristol (MC_UU_00032/1). T.R.-K. was supported by funding from the RCN (number 274611). E.Y. was supported by funding from the RCN (numbers 288083 and 336078) and the European Union (grant agreement numbers 101045526 and 818425). P.C. has received funding from the

University of Oulu and the Academy of Finland Profi6 (decision number AF 336449). S. Seibert acknowledges funding support by the Horizon Europe framework programme (STAGE grant number 101137146, IHEN grant number 101137317, OBELISK grant number 101080465, OBCT grant number 101080250 and TRIGGER grant number 101057739), the Horizon 2020 framework programme (LongITools grant number 874739, EarlyCause grant number 848158) and the research council of Finland (grant number 356888). M.-R.J. acknowledges funding support by the Horizon 2020 framework programme (LongITools grant number 874739, EarlyCause grant number 848158) and by the MRC Centre for Environment and Health funded by the Medical Research Council, UK (MR/S019669/1). D.I.B. acknowledges the Royal Netherlands Academy of Science Professor Award (PAH/6635). A. Ronald received funding from the Simons Foundation (award identifier 724306). K. Rimfeld is supported by a Sir Henry Wellcome Postdoctoral Fellowship (213514/Z/18/Z). N.V.-T. is supported by the Spanish Ministry of Science and Innovation–State Research Agency grant RYC2022-038136-I, cofunded by the European Union FSE+, and grant PID2022-143106OA-I00, cofunded by the European Union FEDER. In addition, N.V.-T. is supported by the William H. Gates Sr. Fellowship from the Alzheimer's Disease Data Initiative. Twins Early Development Study (TEDS) is supported by a programme grant to R.P. from the UK Medical Research Council (grant number MR/VO12878/1 and previously MR/MO21475/1), with additional support from the US National Institutes of Health (grant number AGO46938). The research leading to these results has also received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) (grant agreement number 602768). V.W. is supported by funding from the Simons Foundation for Autism Research Initiative, the Wellcome Trust (214322/Z/18/Z), Horizon-Europe R2D2-MH (grant agreement number 101057385) and UKRI (10063472). J.K.B. has been supported by the EU-AIMS (European Autism Interventions) and AIMS-2-TRIALS programmes, which receive support from Innovative Medicines Initiative Joint Undertaking grant numbers 115300 and 777394, the resources of which are composed of financial contributions from the European Union's FP7 and Horizon 2020 Programmes and from the European Federation of Pharmaceutical Industries and Associations (EFPIA) companies' in-kind contributions, and by AUTISM SPEAKS, Autistica and SFARI; and by the Horizon 2020-supported programme CANDY (grant number 847818). The funders had no role in the design of the study, in the collection, analyses or interpretation of data, in the writing of the paper, or in the decision to publish the results. This publication is the work of the authors, who serve as the guarantors for the contents of this paper. Views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union, the European Research Council Executive Agency or the other funders. Neither the European Union nor the granting authority can be held responsible for them. Open access funding provided by Max Planck Society.

Competing interests

H.L. reports receiving grants from Shire Pharmaceuticals; personal fees from and serving as a speaker for Medice, Shire/Takeda Pharmaceuticals and Evolan Pharma AB; all outside the submitted

work. H.L. is editor-in-chief of *JCPP Advances*. J.A.R.-Q. was on the speakers' bureau and/or acted as a consultant for Biogen, Idorsia, Casen-Recordati, Janssen-Cilag, Novartis, Takeda, Bial, Sincrolab, Neuraxpharm, Novartis, BMS, Medice, Rubió, Uriach, Technofarma and Raffo in the last 3 years. J.A.R.-Q. also received travel awards (air tickets and hotel) for taking part in psychiatric meetings from Idorsia, Janssen-Cilag, Rubió, Takeda, Bial and Medice. The Department of Psychiatry, chaired by J.A.R.-Q., received unrestricted educational and research support from the following companies in the last 3 years: Exeltis, Idorsia, Janssen-Cilag, Neuraxpharm, Oryzon, Roche, Probitas and Rubió. E.D.R. has served as a speaker for Shire Sweden AB, a Takeda Pharmaceutical Company, outside of this work. S.B. discloses that in the last 3 years, he has acted as a consultant or lecturer for Medice, Takeda and LinusBio. S.B. receives royalties for textbooks and diagnostic tools from Hogrefe, UTB, Ernst Reinhardt, Kohlhammer and Liber. S.B. is a partner in NeuroSupportSolutions International AB. The other authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41562-026-02551-z>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41562-026-02551-z>.

Correspondence and requests for materials should be addressed to Beate St Pourcain.

Peer review information *Nature Human Behaviour* thanks Danielle Dick, Nicholas Martin and Felix Trops for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2026

Lucía de Hoyos¹, Fenja Schlag¹, Sanjeevan Jahagirdar¹, Elizabeth C. Corfield^{2,3,4}, Andrea G. Allegrini^{5,6}, Danielle Admiraal¹, Eveline L. de Zeeuw⁷, Ilja M. Nolte⁸, Natalia Llonga^{9,10,11,12}, Alexander Neumann¹³, Katherine Lange^{14,15}, Simone van den Bedem¹, Ebba Du Rietz¹⁶, Ehsan Motazed^{17,18}, Else Eising¹, Nicole Y. T. Ng¹, Teemu Palviainen¹⁹, Carol A. Wang^{20,21}, Elisabeth Thiering^{22,23}, Sofia Scatolin^{1,24}, Priyanka Choudhary²⁵, Natalia Vilor-Tejedor^{26,27,28,29}, Zhijie Liao³⁰, Silvia Alemany^{9,11,31}, Casper-Emil T. Pedersen³², Nora Fernandez-Jimenez^{33,34}, Megan L. Campbell^{35,36}, Polina Girchenko³⁷, Eli Barthome³⁸, L. Caitlin Martin³⁸, Hanna Seelemeyer^{39,40}, Afsheen Kumar^{39,40}, Alexandre Jeanne^{39,40}, Tarunveer S. Ahluwalia^{32,41,42,43}

Melissa H. Black^{44,45,46}, Jan K. Buitelaar⁴⁷, Simon E. Fisher^{1,48}, Anni Heiskala²⁵, María Hernández-Lorca^{32,49}, Vincent W. Jaddoe^{50,51}, Marjo-Riitta Järvelin^{25,52,53,54}, Sheri-Michelle Koopowitz³⁵, Marilyn T. Lake^{35,36}, Paul Lichtenstein¹⁶, Sabrina Llop^{55,56}, Mannan Luo^{13,50}, Anni Malmberg³⁷, Sergi Marí^{33,34}, Albertine J. Oldehinkel⁵⁷, Zdenka Pausova^{58,59}, Craig E. Pennell^{20,21}, Robert Plomin⁶, Ted Reichborn-Kjennerud^{60,61}, Sheena Reilly^{15,62,63}, Alina Rodriguez^{64,65,66,67}, Thomas Rolland^{38,68}, Angelica Ronald⁶⁹, Katri Räikkönen³⁷, Jean Shin⁵⁹, Maria Soler Artigas^{9,10,11,12}, Ashley E. Tate¹⁶, Henning Tiemeier^{70,71}, Ellen Verhoef¹, Eero Vuoksima¹⁹, Melissa Wake^{14,15}, Varun Warriar^{72,73}, Eivind Ystrom^{2,74,75}, Sven Bölte^{45,46,76}, Christine Ecker^{39,40,77}, Thomas Bourgeron³⁸, Jari Lahti³⁷, Dan J. Stein⁷⁸, Jose Ramon Bilbao^{33,34,79}, Klaus Bønnelykke³², Mariona Bustamante^{31,56,80}, Tomas Paus⁸¹, Christel M. Middeldorp^{82,83,84,85,86}, Jordi Sunyer^{29,31,56,80}, Sylvain Sebert²⁵, Marie Standl^{22,87}, Andrew J. O. Whitehouse⁸⁸, Jaakko Kaprio¹⁹, J. Antoni Ramos-Quiroga^{9,10,11,89}, Marta Ribases^{9,10,11,12}, Miguel Casas^{89,90,91,92}, Rosa Bosch^{11,90,93,94}, Angela T. Morgan^{14,95}, Tanja G. M. Vrijkotte^{17,18}, Henrik Larsson^{16,96}, Charlotte A. M. Cecil^{13,97}, Catharina A. Hartman⁵⁷, Dorret I. Boomsma⁹⁸, Kaili Rimfeld^{6,99}, Alexandra Havdahl^{2,3,74} & Beate St Pourcain^{1,4,48}✉

¹Language and Genetics Department, Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands. ²PsychGen Centre for Genetic Epidemiology and Mental Health, Norwegian Institute of Public Health, Oslo, Norway. ³Psychiatric Genetic Epidemiology Group, Research Department, Lovisenberg Diaconal Hospital, Oslo, Norway. ⁴MRC Integrative Epidemiology Unit, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK. ⁵Department of Clinical, Educational and Health Psychology, Division of Psychology and Language Sciences, University College London, London, UK. ⁶Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK. ⁷Department of Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. ⁸Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. ⁹Psychiatric Genetics Unit, Group of Psychiatry, Mental Health and Addiction, Vall d'Hebron Research Institute (VHIR), Universitat Autònoma de Barcelona, Barcelona, Spain. ¹⁰Department of Mental Health, Hospital Universitari Vall d'Hebron, Barcelona, Spain. ¹¹Biomedical Network Research Centre on Mental Health (CIBERSAM), Madrid, Spain. ¹²Department of Genetics, Microbiology, and Statistics, Faculty of Biology, Universitat de Barcelona (UB), Barcelona, Spain. ¹³Department of Child and Adolescent Psychiatry/Psychology, Erasmus MC University Medical Center Rotterdam, Rotterdam, the Netherlands. ¹⁴Murdoch Children's Research Institute, Royal Children's Hospital, Parkville, Victoria, Australia. ¹⁵Department of Paediatrics, University of Melbourne, Parkville, Victoria, Australia. ¹⁶Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. ¹⁷Department of Public and Occupational Health, Amsterdam UMC, University of Amsterdam, Amsterdam, the Netherlands. ¹⁸Amsterdam Public Health Research Institute, Amsterdam, the Netherlands. ¹⁹Institute for Molecular Medicine Finland (FIMM), HiLIFE, University of Helsinki, Helsinki, Finland. ²⁰School of Medicine and Public Health, University of Newcastle, Newcastle, New South Wales, Australia. ²¹Mothers and Babies Research Program, Hunter Medical Research Institute, New Lambton Heights, New South Wales, Australia. ²²Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. ²³Division of Metabolic and Nutritional Medicine, Department of Paediatrics, Dr. von Hauner Children's Hospital, LMU University Hospital, LMU Munich, Munich, Germany. ²⁴Jacobs Center for Productive Youth Development, University of Zurich, Zurich, Switzerland. ²⁵Research Unit of Population Health, Faculty of Medicine, University of Oulu, Oulu, Finland. ²⁶BarcelonaBeta Brain Research Center, Pasqual Maragall Foundation, Barcelona, Spain. ²⁷Department of Human Genetics, Radboud University Medical Center, Nijmegen, the Netherlands. ²⁸Centre for Genomic Regulation (CRG), The Barcelona Institute for Science and Technology, Barcelona, Spain. ²⁹Hospital del Mar Medical Research Institute, Barcelona, Spain. ³⁰Centre de Recherche du CHU Sainte-Justine, Montreal, Quebec, Canada. ³¹Institute for Global Health (ISGlobal), Barcelona, Spain. ³²Copenhagen Prospective Studies on Asthma in Childhood (COPSAC), Herlev and Gentofte Hospital, University of Copenhagen, Copenhagen, Denmark. ³³Department of Genetics, Physical Anthropology and Animal Physiology, University of the Basque Country (UPV/EHU), Leioa, Spain. ³⁴Biobizkaia Health Research Institute, Barakaldo, Spain. ³⁵Department of Psychiatry and Neuroscience Institute, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa. ³⁶Child and Adolescent Health SAMRC Unit, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa. ³⁷Department of Psychology, Faculty of Medicine, University of Helsinki, Helsinki, Finland. ³⁸Human Genetics and Cognitive Functions, Institut Pasteur, CNRS UMR3571, Institut Universitaire de France, Université Paris Cité, Paris, France. ³⁹Department of Child and Adolescent Psychiatry, University Hospital, Goethe University Frankfurt, Frankfurt am Main, Germany. ⁴⁰Cooperative Brain Imaging Center (CoBIC), Goethe University Frankfurt, Frankfurt am Main, Germany. ⁴¹Steno Diabetes Center Copenhagen, Herlev, Denmark. ⁴²The Bioinformatics Center, Department of Biology, University of Copenhagen, Copenhagen, Denmark. ⁴³Department of Science and Environment, Roskilde University, Roskilde, Denmark. ⁴⁴Department of Community and Clinical Health, School of Allied Health, Human Services and Sport, La Trobe University, Melbourne, Victoria, Australia. ⁴⁵Center of Neurodevelopmental Disorders (KIND), Department of Women's and Children's Health, Centre for Psychiatry Research, Karolinska Institutet and Region Stockholm, Stockholm, Sweden. ⁴⁶Curtin School of Allied Health, Curtin University, Perth, Western Australia, Australia. ⁴⁷Karakter Child and Adolescent Psychiatry, Nijmegen, the Netherlands. ⁴⁸Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, the Netherlands. ⁴⁹CNSR, Mental Health Centre Glostrup, Copenhagen University Hospital, Copenhagen, Denmark. ⁵⁰The Generation R Study Group, Erasmus MC University Medical Center Rotterdam, Rotterdam, the Netherlands. ⁵¹Department of Pediatrics, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands. ⁵²MRC Centre for Environment and Health, Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK. ⁵³Unit of Primary Care, Oulu University Hospital, Oulu, Finland. ⁵⁴Department of Life Sciences, College of Health and Life Sciences, Brunel University London, London, UK. ⁵⁵Epidemiology and Environmental Health Joint Research Unit, FISABIO-Universitat Jaume I-Universitat de València, Valencia, Spain. ⁵⁶Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP), Madrid, Spain. ⁵⁷Interdisciplinary Center Psychopathology and Emotion Regulation, Department of Psychiatry, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. ⁵⁸Department of Pediatrics, University of Montreal, Montreal, Quebec, Canada. ⁵⁹Centre Hospitalier Universitaire Sainte-Justine, Montreal, Quebec, Canada. ⁶⁰PsychGen Centre for Genetic Epidemiology and Mental Health, Oslo, Norway. ⁶¹Institute of Clinical Medicine, University of Oslo, Oslo, Norway. ⁶²Griffith University, Southport, Queensland, Australia. ⁶³Murdoch Children's Research Institute, Parkville, Victoria, Australia. ⁶⁴Anglia Ruskin University, School of Allied Health, Chelmsford, UK. ⁶⁵Agency for Science, Technology and Research

(A*STAR), Institute of Human Development and Potential, Singapore, Singapore. ⁶⁶National University of Singapore, Department of Psychological Medicine, Singapore, Singapore. ⁶⁷Imperial College London, School of Public Health, Department of Epidemiology and Biostatistics, London, UK. ⁶⁸Centre de Recherche en Biologie Cellulaire de Montpellier (CRBM), CNRS UMR5237, Montpellier, France. ⁶⁹School of Psychology, University of Surrey, Guildford, UK. ⁷⁰Department of Social and Behavioral Sciences, Harvard T.H. Chan School of Medicine, Boston, MA, USA. ⁷¹Department of Child and Adolescent Psychiatry, Erasmus University Medical Center, Rotterdam, the Netherlands. ⁷²Department of Psychiatry, University of Cambridge, Cambridge, UK. ⁷³Department of Psychology, University of Cambridge, Cambridge, UK. ⁷⁴PROMENTA Research Centre, Department of Psychology, University of Oslo, Oslo, Norway. ⁷⁵Centre for Research on Equality in Education, Faculty of Educational Sciences, University of Oslo, Oslo, Norway. ⁷⁶Child and Adolescent Psychiatry, Stockholm Health Care Services, Region Stockholm, Stockholm, Sweden. ⁷⁷Department of Forensic and Neurodevelopmental Sciences, Institute of Psychiatry, Psychology, and Neuroscience, King's College London, London, UK. ⁷⁸SAMRC Unit on Risk and Resilience in Mental Disorders, Department of Psychiatry and Neuroscience Institute, University of Cape Town, Cape Town, South Africa. ⁷⁹Center for Biomedical Research in Diabetes and Associated Metabolic Disorders (CIBERDEM), Madrid, Spain. ⁸⁰Universitat Pompeu Fabra (UPF), Barcelona, Spain. ⁸¹Departments of Psychiatry and Neuroscience, Faculty of Medicine and Center Hospitalier Universitaire Sainte-Justine, University of Montreal, Montreal, Quebec, Canada. ⁸²Department of Child and Adolescent Psychiatry and Psychology, Amsterdam UMC, Amsterdam Reproduction and Development research Institute, Amsterdam Public Health Research Institute, Amsterdam, the Netherlands. ⁸³Arkin Mental Health Care, Amsterdam, the Netherlands. ⁸⁴Levvel, Academic Center for Child and Adolescent Psychiatry, Amsterdam, the Netherlands. ⁸⁵Child Health Research Centre, University of Queensland, Brisbane, Queensland, Australia. ⁸⁶Child and Youth Mental Health Service, Children's Health Queensland Hospital and Health Service, Brisbane, Queensland, Australia. ⁸⁷German Center for Child and Adolescent Health (DZKJ), Munich, Germany. ⁸⁸The Kids Research Institute Australia, University of Western Australia, Perth, Western Australia, Australia. ⁸⁹Department of Psychiatry and Forensic Medicine, Universitat Autònoma de Barcelona (UAB), Barcelona, Spain. ⁹⁰SJD MIND Schools Program, Hospital Sant Joan de Déu, Institut de Recerca Sant Joan de Déu, Esplugues de Llobregat, Spain. ⁹¹Fundació d'Investigació Salut i Progrés, Barcelona, Spain. ⁹²Instituto para el Desarrollo de Terapias Avanzadas en Neurociencias, Barcelona, Spain. ⁹³Divisió de Salut Mental, Althaia Xarxa Assistencial Universitària de Manresa, Manresa, Spain. ⁹⁴Grup de Recerca en Innovació en Salut Mental i Benestar Emocional (ISaMBeS), Institut de Recerca i Innovació en Ciències de la Vida i de la Salut a la Catalunya Central (IRIS-CC), Vic, Spain. ⁹⁵Speech Pathology, School of Health Sciences, University of Melbourne, Parkville, Melbourne, Victoria, Australia. ⁹⁶School of Medical Sciences, Örebro University, Örebro, Sweden. ⁹⁷Department of Epidemiology, Erasmus MC, Rotterdam, the Netherlands. ⁹⁸Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. ⁹⁹Department of Psychology, Royal Holloway University of London, London, UK. ✉ e-mail: beate.stpourcain@mpi.nl

Extended Data Table 1 | Overview of participating cohorts

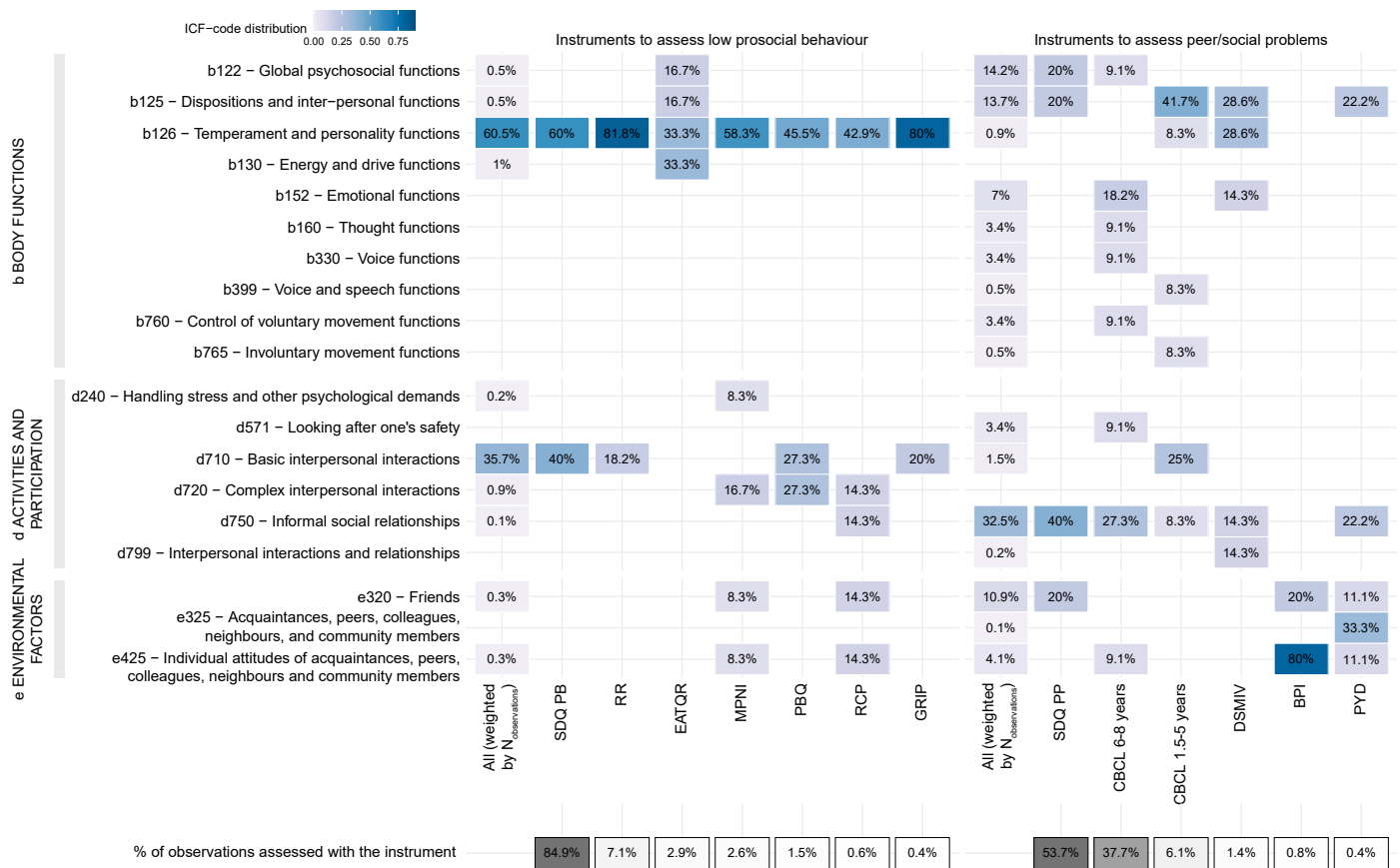
Abbreviation	Cohort name	Country	Domain	Reporters	Median age (in years)	N _{ind}	N _{obs}
ABCD-NL	Amsterdam Born Children and their Development (ABCD)	Netherlands	LPB PSP	P, T, S	5,12,16	1,133	10,475
ABCD-US	Adolescent Brain Cognitive Development (ABCD) Study	United States of America	LPB PSP	P, S	10,11,12,13	7,691	68,594
ALSPAC	Avon Longitudinal Study of Parents and Children (ALSPAC)	United Kingdom	LPB PSP	P, T, S	4,7,8,10,11,12,13,17,25	6,174	91,905
BREATHE	BRain dEvelopment and Air polluTion ultrafine particles in scHool childrEn (BREATHE)	Spain	LPB PSP	P	8	1,616	3,231
CATSS	Child and Adolescent Twin Study in Sweden (CATSS)	Sweden	LPB	P, S	15,16	5,396	10,552
COPSAC	Copenhagen Prospective Studies on Asthma in Childhood 2010 (COPSAC2010)	Denmark	LPB PSP	P	6	553	1,106
ELVS	Early Language in Victoria Study (ELVS)	Australia	LPB PSP	P, T, S	4,5,6,7,9,11,13	641	10,100
FinnTwin12	FinnTwin12	Finland	LPB	P, T, S	12,14,18	1,399	6,413
GenR	Generation R (GenR)	Netherlands	LPB PSP	P, T, S	2,3,6,10	2,038	16,183
GINILISA	GINI/LISA Birth Cohort Studies	Germany	LPB PSP	P, S	10,15	1,369	5,113
GLAKU	GLAKU birth cohort	Finland	PSP	P	12	309	309
INMA_GSA	INfancia y Medio Ambiente - Environment and Childhood (INMA) study - Gipuzcoa	Spain	LPB PSP	P	4,8	312	1,046
INMA_Omni1	INfancia y Medio Ambiente - Environment and Childhood (INMA) study - Sabadell and Valencia	Spain	LPB PSP	P	7	584	1,167
INSchoolw1	INSchool wave1	Spain	LPB PSP	P, T, S	10,13	819	4,709
INSchoolw2	INSchool wave2	Spain	LPB PSP	P, T, S	10,13	2,839	16,726
INSchoolw3	INSchool wave3	Spain	LPB PSP	P, T, S	13	960	5,729
LSAC	Longitudinal Study of Australian Children study (LSAC)	Australia	LPB PSP	P	5,7,9,11,13	1,372	11,965
MoBa	Norwegian Mother, Father and Child Cohort Study (MoBa)	Norway	LPB	P	3	16,768	16,768
MUSP	Mater University Study of Pregnancy (MUSP)	Australia	PSP	P, S	14	1,226	2,451
NFBC1986	North Finish Birth Cohort 1986 (NFBC1986)	Finland	PSP	Self	16	3,310	3,310
NTR	Netherlands Twin Register (NTR)	Netherlands	PSP	P, T, S	3,6,7,10,12,15,17	4,535	40,461
The Raine Study	The Raine Study	Australia	PSP	P, S	2,8,11*,14	1,366	6,332
SYS	Saguenay Youth Study (SYS)	Canada	LPB PSP	S	14	1,006	1,931
TEDS	Twins Early Development Study (TEDS)	United Kingdom	LPB PSP	P, T, S	2,3,4,7,9,11,12,16	8,243	126,817
TRAILS	Tracking Adolescents' Individual Lives Survey (TRAILS)	Netherlands	LPB PSP	P, T, S	11,13,16,19,22,26,29	1,662	27,853

Table describing the domains, reporters and ages provided by each cohort. Further details on the individual measures included are available in Supplementary Table 1. * The follow-up in The Raine Study was conducted at 10 years, but we report 11 years as we use the rounded median age across all GWAS; in this case, the median age for the follow-up at 10 years was 10.58 (s.d. 0.2) years. Abbreviations: LPB (low prosocial behavior), N_{ind} (number of individuals), N_{obs} (number of observations), P (parent-reported), PSP (peer/social problems), T (teacher-reported), S (self-reported).

Extended Data Table 2 | Meta-regression coefficients for social behaviour lead SNPs across social domain, reporter, and age

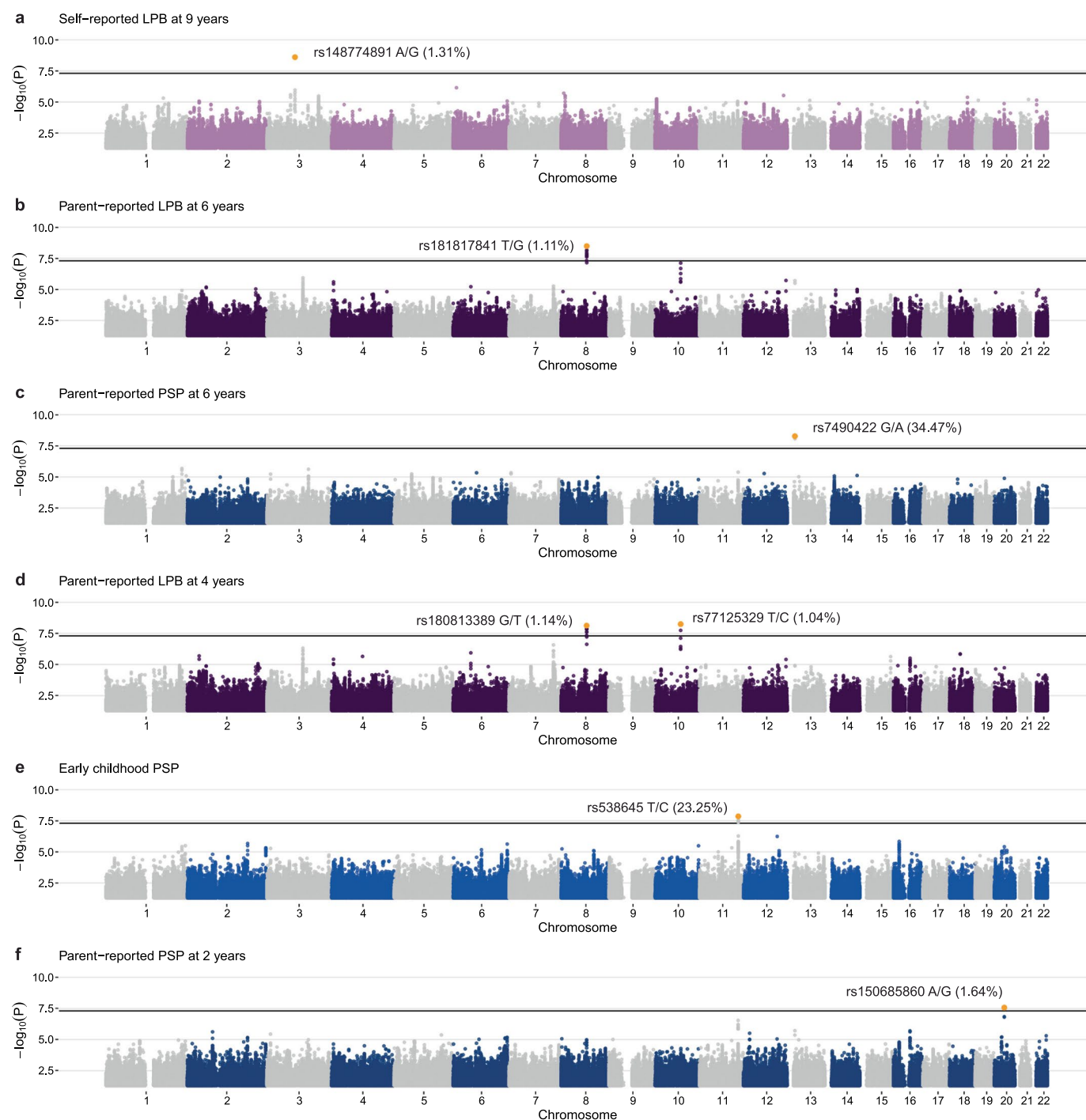
rsID	chr	position (hg19)	Intercept (θ_0)	Domain-PSP (θ_{PSP})	Reporter-teacher ($\theta_{Teacher}$)	Reporter-self (θ_{Self})	Age (θ_{Age})	Age square (θ_{Age^2})	N_{ind}	N_{obs}	$N_{cohorts}$	N_{GWAS}
rs148774891	3	86059243	0.026 s.e.=0.026 p=0.32	-0.031 s.e.=0.028 p=0.255	-0.0082 s.e.=0.0335 p=0.807	0.17 s.e.=0.032 p=1.25x10 ⁻⁷	-0.005 s.e.=0.003 p=0.105	<0.001 s.e.<0.001 p=0.311	50,442	430,165	18	162
rs181817841	8	78481590	0.111 s.e.=0.025 p=5.49x10 ⁻⁶	-0.054 s.e.=0.026 p=0.0389	-0.087 s.e.=0.032 p=5.97x10 ⁻³	-0.086 s.e.=0.031 p=5.72x10 ⁻³	-0.010 s.e.=0.003 p=1.32x10 ⁻³	<0.001 s.e.<0.001 p=0.650	52,740	430,676	20	170
rs77125329	10	78297483	-0.083 s.e.=0.024 p=5.24x10 ⁻⁴	0.100 s.e.=0.025 p=7.67x10 ⁻⁵	0.019 s.e.=0.032 p=0.546	0.023 s.e.=0.029 p=0.439	0.007 s.e.=0.003 p=9.72x10 ⁻³	-0.001 s.e.<0.001 p=0.0130	71,002	462,516	22	180
rs538645	11	118711069	0.002 s.e.=0.006 p=0.709	-0.013 s.e.=0.007 p=0.0488	-0.012 s.e.=0.008 p=0.130	-0.007 s.e.=0.007 p=0.319	0.003 s.e.=0.001 p=1.04x10 ⁻⁵	<0.001 s.e.<0.001 p=0.0251	56,300	453,163	24	194
rs7490422	13	23284225	-0.024 s.e.=0.006 p=4.65x10 ⁻⁵	-0.006 s.e.=0.006 p=0.302	0.015 s.e.=0.008 p=0.0546	0.018 s.e.=0.007 p=0.0113	0.001 s.e.=0.001 p=0.177	<0.001 s.e.<0.001 p=0.431	53,458	447,481	22	190
rs150685860	20	30561451	-0.002 s.e.=0.020 p=0.924	-0.036 s.e.=0.021 p=0.0870	0.006 s.e.=0.025 p=0.819	0.086 s.e.=0.025 p=6.74x10 ⁻⁴	0.009 s.e.=0.002 p=2.73x10 ⁻⁴	-0.001 s.e.<0.001 p=2.21x10 ⁻⁴	43,592	374,538	20	176

Mixed-effects meta-regression was used to meta-analyse observed genetic effects of social behaviour across social domains, reporter, and age. Fixed-effects coefficients are shown for the lead SNPs identified in this meta-analysis. Effects are reported for the intercept (θ_0), and the moderators included in the meta-regression: social domain (θ_{PSP}), reporter ($\theta_{Teacher}$, θ_{Self}), and age (θ_{Age} , θ_{Age^2}). The fixed-effect intercept (θ_0) represents the parent-reported LPB SNP effect at the N-weighted grand mean age (10.14 years). The fixed-effect domain coefficient (θ_{PSP}) represents the different in SNP effects for PSP domain relative to LPB domain (reference category). The fixed-effect reporter coefficients ($\theta_{Teacher}$, θ_{Self}) represent the difference in SNP effects for teacher and self-reports, respectively, relative to parent reports (reference category). The fixed-effect age coefficients (θ_{Age} , θ_{Age^2}) represent the linear and quadratic age-related moderation, respectively, capturing differences in SNP effects for an increase in one year of age. Association statistics were derived using mixed-effects meta-regression models with two-sided tests. A full table including meta-regression coefficients for all SNPs passing the genome-wide significance threshold ($p < 5 \times 10^{-8}$) is available at Supplementary Table 5. Abbreviations: N_{ind} (number of individuals), N_{GWAS} (number of sets of GWAS summary statistics), N_{obs} (number of observations).



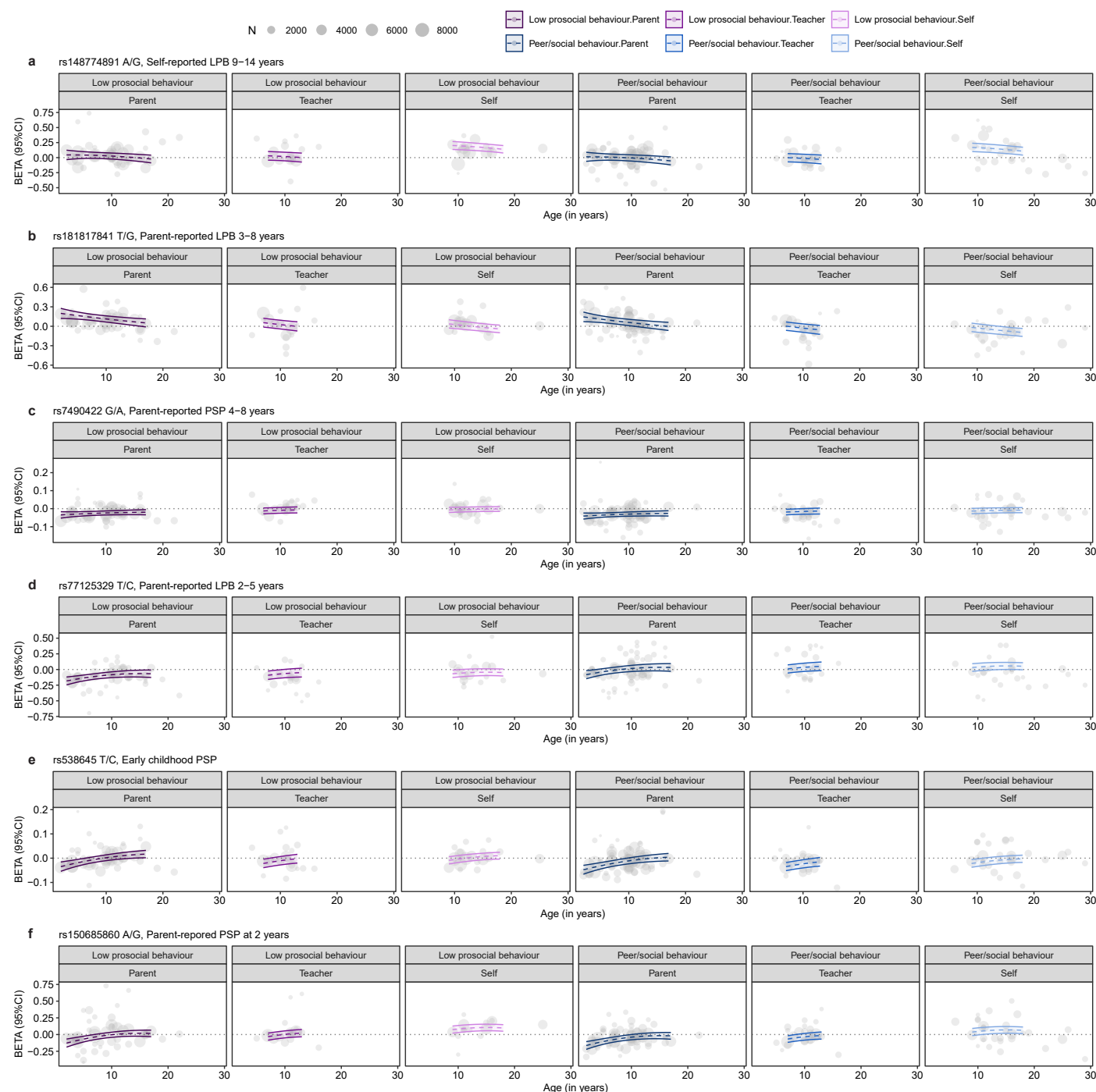
Extended Data Fig. 1 | ICF alignment of social behaviour measures. The y-axis shows the different codes from the ICF-CY grouped according to the ICF category. The x-axis shows for low prosocial behaviour (left) and peer/social problems (right) the distribution of the ICF codes across the questionnaires used to assess each social domain, as well as the overall contribution weighted by the number of observations ($N_{\text{observations}}$). Note that different versions of the CBCL were included assessing social problems from 6 years onwards, and these were all included within the instrument CBCL 6–8 years (see Supplementary Note 2 for a detailed description). An overview of item classification is available at Supplementary Table 3. Abbreviations: BPI (Berkeley Puppet Interview), CBCL

(Child Behaviour Checklist), DSMIV (Diagnostic and statistical manual of mental disorders IV), EATQR (Early Adolescent Temperament Questionnaire-Revised), GRIP (Groupe de Recherche sur l'Inadaptation Psychosociale), ICF (WHO's International Classification of Functioning, Disability and Health, Children & Youth Version), MPNI (Multidimensional Peer Nomination Inventory), PB (Prosocial Behaviour), PBQ (Prosocial Behaviour Questionnaire), PP (Peer Problems), PYD (Positive Youth Development), RCP (Revised Class Play method of peer assessment), RR (Revised Rutter Parent Scale for Preschool Children), SDQ (Strengths and Difficulties Questionnaire).



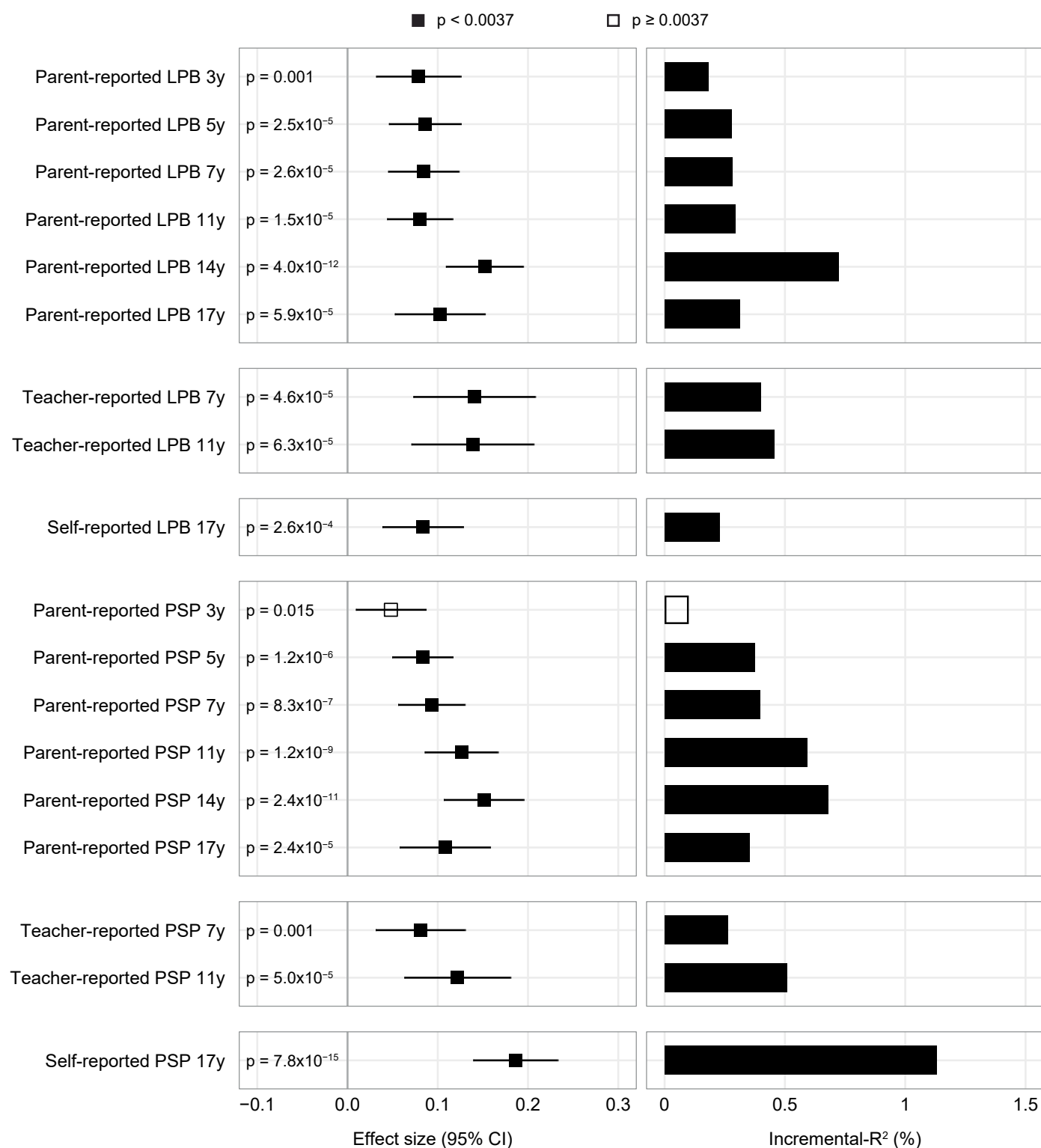
Extended Data Fig. 2 | Manhattan plots. Corresponding to the six identified independent GWA loci (Table 1), Manhattan plots for the GWAS with the strongest signal are shown (a-f). Each dot represents a SNP. The $-\log_{10}(p)$ is shown on the y-axis, while the x-axis indicates its genomic coordinates. Genetic effect estimates were derived using mixed-effects meta-regression models

with two-sided tests. A genome-wide significance threshold was set to 5×10^{-8} (0.05/1,000,000) and indicated by the solid horizontal black line. For each GWAS, the lead SNPs at each locus passing this genome-wide significance threshold have been annotated with a yellow dot along with their alleles (effect/other allele) and their minor allele frequency.



Extended Data Fig. 3 | Derived meta-analysed effect sizes for genome-wide associated loci across social domains, reporters, and ages. Corresponding to the six identified independent GWA loci (Table 1), we report meta-analysed effect sizes for each index SNP across 66 low-resolution GWAS (a–f). Dashed lines represent the meta-analysed SNP effects, and shaded areas indicate the associated 95% confidence intervals. For each SNP, genetic effects from up to 195

GWAS were meta-analysed using mixed-effects meta-regression (see Methods for a detailed description). Grey points represent SNP effect size estimates from individual cohort-level GWAS contributing to the meta-analysis (up to 195 GWAS estimates). The point size is proportional to the sample size (N_{obs}). The meta-analysed SNP effects are presented across age, stratified by social domain and reporter.



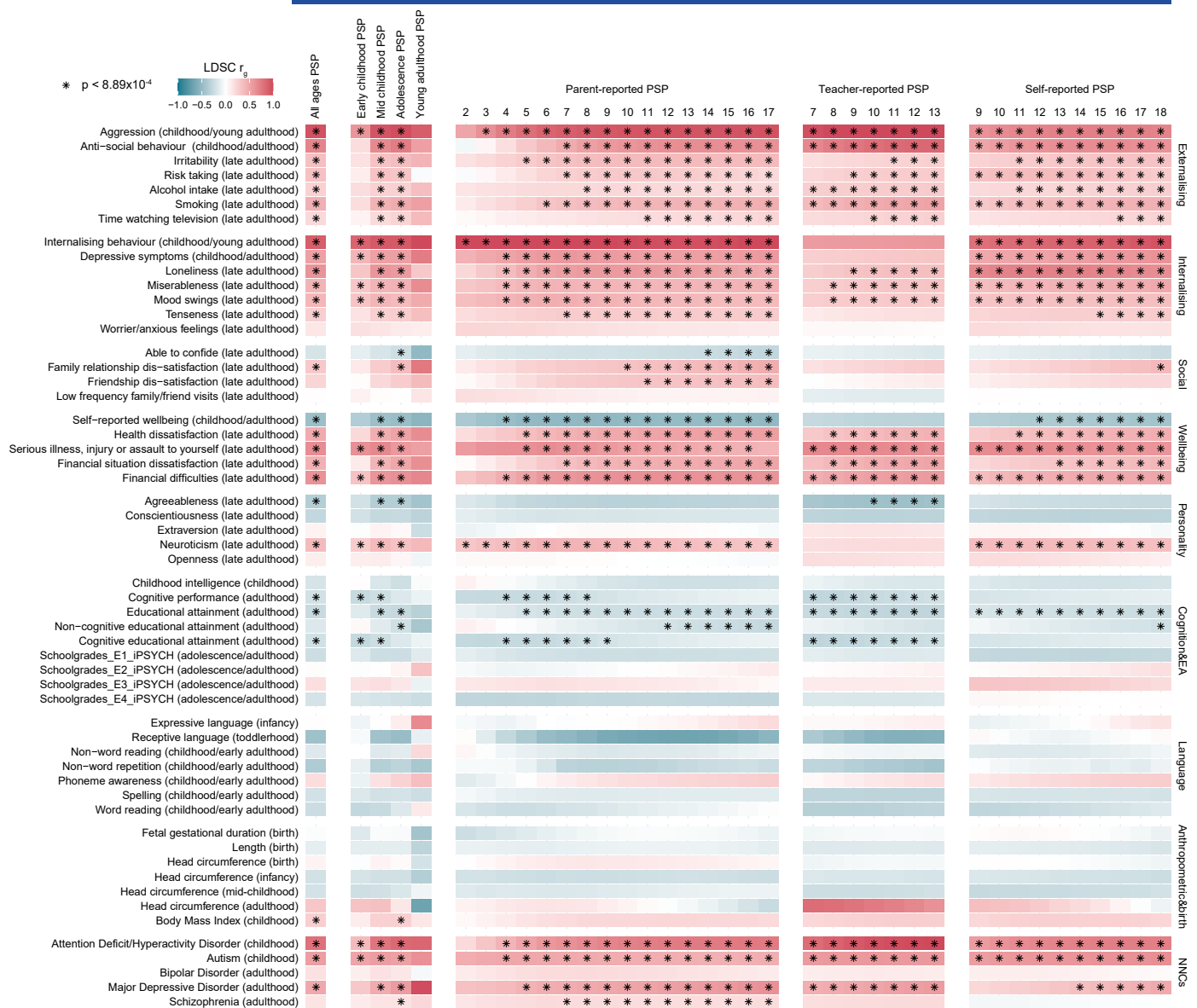
Extended Data Fig. 4 | Validation analysis of fully-matched social behaviour polygenic scores in the Millennium Cohort Study. Polygenic scores were computed by matching discovery GWAS to target phenotypes (from the Millennium Cohort Study) according to social domains, reporters and age. In the left panel, squares represent estimated polygenic score effect sizes, and horizontal error bars indicate 95% confidence intervals. In the right panel, the amount of variance explained is expressed as incremental-R² (%).

Associations passing the multiple-testing threshold ($p < 0.0037$) are shown with a filled rectangle, and with an empty rectangle otherwise. Sample sizes for discovery social behaviour GWAS are presented in Supplementary Table 4. Sample sizes for target phenotypes are presented in Supplementary Table 15. Estimates, along with their standard errors and incremental-R², are reported in Supplementary Table 16. Abbreviations: LPB (Low prosocial behaviour), PSP (Peer and social problems).

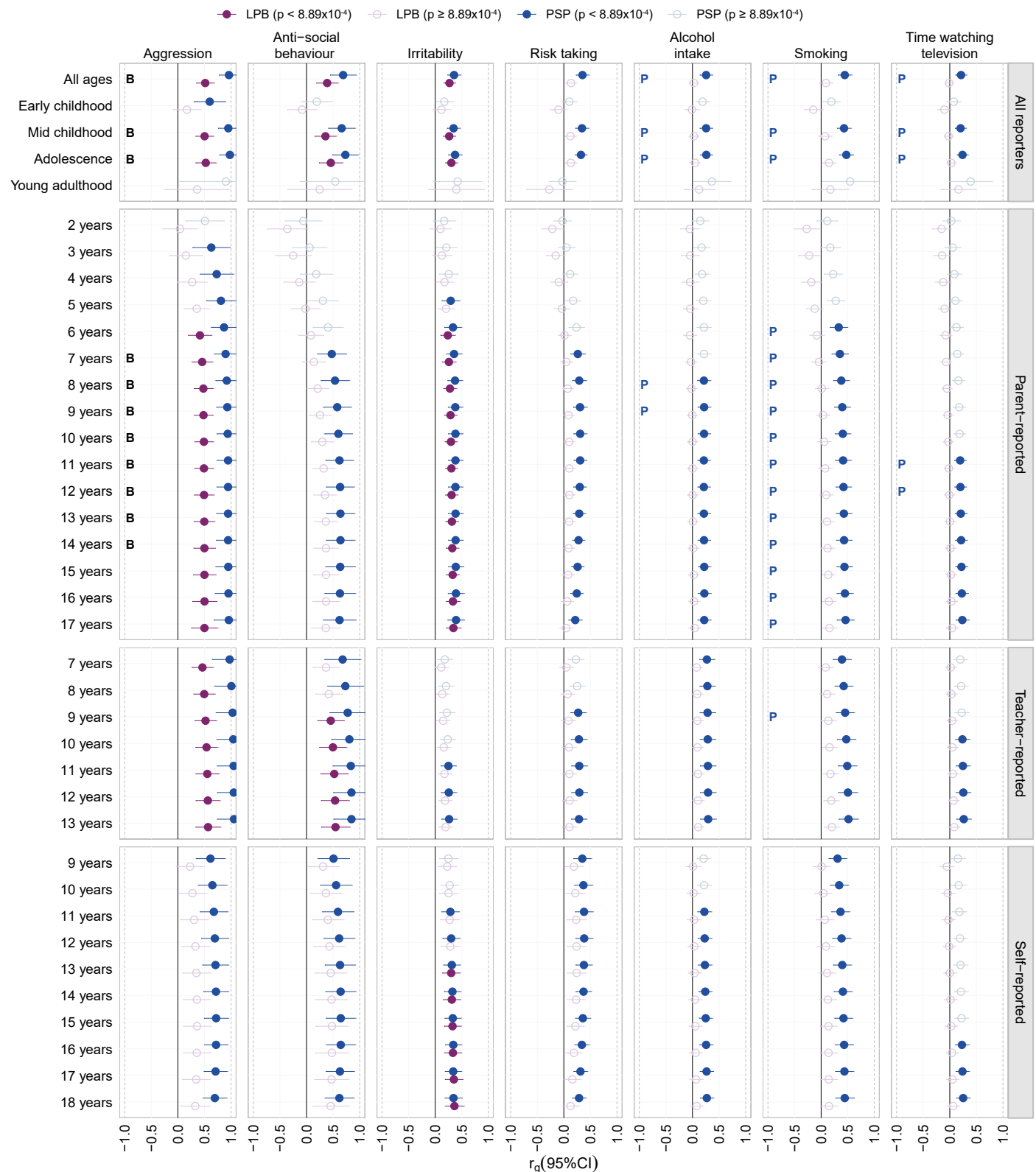


Extended Data Fig. 5 | Genetic correlations of LPB with population-based traits and NNCs. Linkage disequilibrium score (LDSC) genetic correlation estimates with low-resolution, medium-resolution and high-resolution LPB GWAS are shown as a heatmap. Genetic correlations passing the multiple-testing threshold ($p < 8.89 \times 10^{-4}$, two-sided Wald tests) are marked with an asterisk.

Genetic correlation estimates, along with their standard errors and p-values, are reported in Supplementary Table 24. Sample sizes for LPB GWAS are presented in Supplementary Table 4. GWAS sample sizes for population-based traits and NNCs are presented in Supplementary Table 22. Abbreviations: LPB (Low prosocial behaviour), NNCs (Neurodevelopmental and neuropsychiatric conditions).

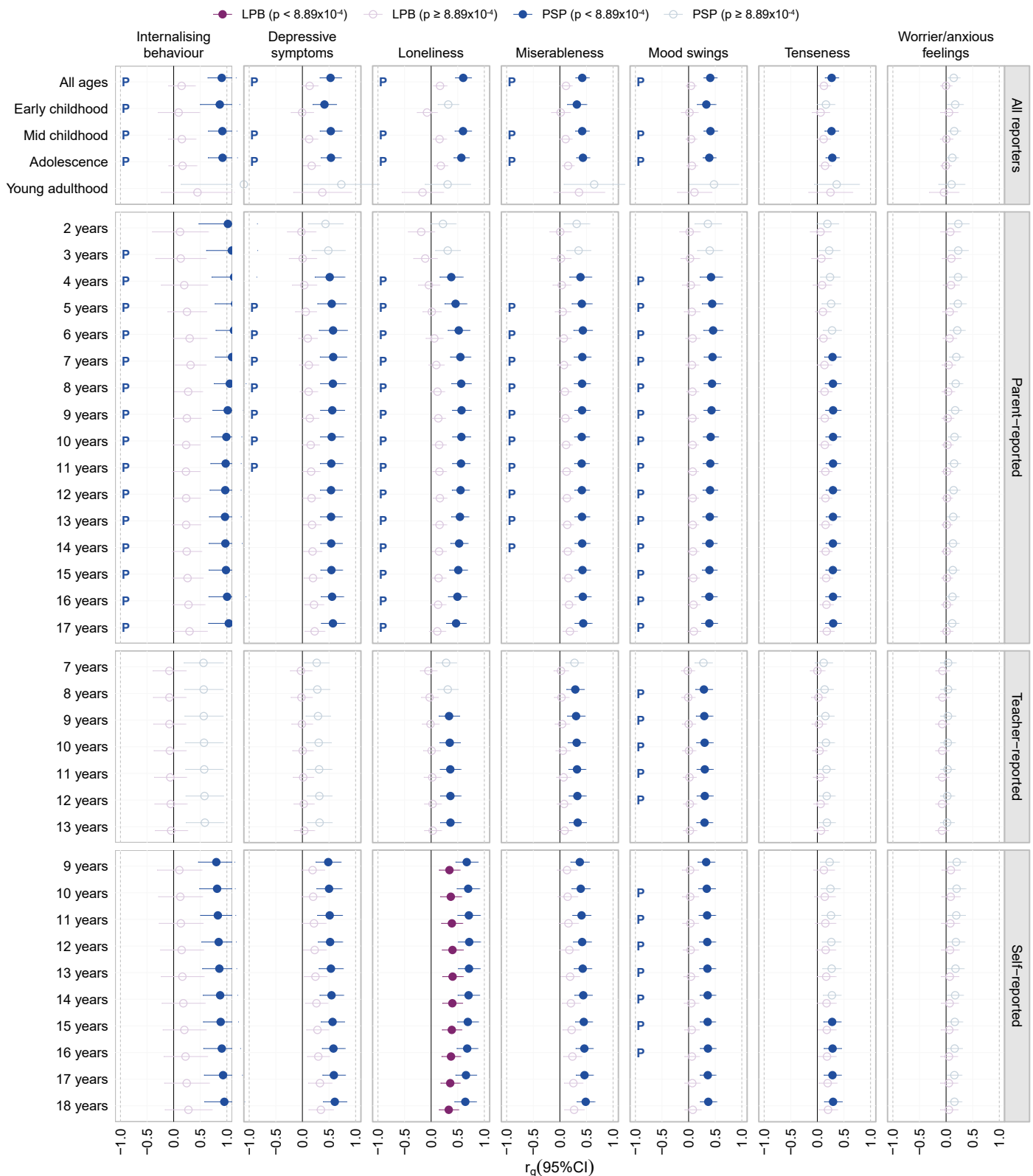


Genetic correlation estimates, along with their standard errors and p-values, are reported in Supplementary Table 24. Sample sizes for PSP GWAS are presented in Supplementary Table 4. GWAS sample sizes for population-based traits and NNCs are presented in Supplementary Table 22. Abbreviations: PSP (Peer and social problems), NNCs (Neurodevelopmental and neuropsychiatric conditions).



Extended Data Fig. 7 | Contextual genetic correlations of social behaviour with externalizing behaviour phenotypes. LDSC genetic correlation estimates (r_g) with low-resolution, medium-resolution and high-resolution LPB and PSP GWAS are shown along with their 95% confidence intervals. Filled circles indicate genetic correlation estimates passing the multiple-testing threshold ($p < 8.89 \times 10^{-4}$, two-sided Wald tests). r_g estimates, standard errors, and p-values are reported in Supplementary Table 24. Differences between PSP and LPB

genetic correlations at different moderator levels, consistent with trait-specific effects indicated by non-overlapping 95% confidence bands, are shown for PSP correlations (blue P), LPB correlations (purple L) or both (black B), once they pass the multiple-testing threshold. Sample sizes for social behaviour GWAS are presented in Supplementary Table 4. GWAS sample sizes for externalizing traits are presented in Supplementary Table 22.



Extended Data Fig. 8 | Contextual genetic correlations of social behaviour with internalising behaviour phenotypes. LDSC genetic correlation estimates (r_g) with low-resolution, medium-resolution and high-resolution LPB and PSP GWAS are shown along with their 95% confidence intervals. Filled circles indicate genetic correlation estimates passing the multiple-testing threshold ($p < 8.89 \times 10^{-4}$, two-sided Wald tests). r_g estimates, standard errors, and p-values are reported in Supplementary Table 24. Differences between PSP and LPB

genetic correlations at different moderator levels, consistent with trait-specific effects indicated by non-overlapping 95% confidence bands, are shown for PSP correlations (blue P), LPB correlations (purple L) or both (black B), once they pass the multiple-testing threshold. r_g estimates, standard errors, and p-values are reported in Supplementary Table 24. Sample sizes for social behaviour GWAS are presented in Supplementary Table 4. GWAS sample sizes for externalising traits are presented in Supplementary Table 22.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | | |
|--------------------------|--|
| n/a | Confirmed |
| ☐ | ☒ The exact sample size (<i>n</i>) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| ☐ | ☒ A description of all covariates tested |
| ☐ | ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted
<i>Give P values as exact values whenever suitable.</i> |
| ☐ | ☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☐ | ☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☐ | ☒ Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i>), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	No software was used for data collection for the purpose of this study.
Data analysis	Code used in this study is available at https://gitlab.gwdg.de/pghc/social-behaviour/social-behaviour-gwas-2026. To conduct data analysis, this study used openly available software and codes. Specifically: abagen toolbox (v0.1.1, https://github.com/rmarkello/abagen) FUMA (v1.6.0, https://fuma.ctglab.nl/) GCTA-COJO (v1.94, https://yanglab.westlake.edu.cn/software/gcta/) GEMMA (https://github.com/genetics-statistics/GEMMA) multivariate GWAMA (v1.2.6, https://github.com/baselmans/multivariate_GWAMA) LDSC (v1.0.1, https://github.com/bulik/ldsc) METAL (v2011-03-25, https://csg.sph.umich.edu/abecasis/metal/) MAGMA (v1.08 and v1.10, https://ctg.cncr.nl/software/magma) PLINK (v1.9, https://www.cog-genomics.org/plink/1.9/) ProbABEL (https://github.com/GenABEL-Project/ProbABEL) PRScs (Jun-04-2021, https://github.com/getian107/PRScs) PRScsx (Nov-21-2024, https://github.com/getian107/PRScsx) SNPTEST (v2.4.1, https://mathgen.stats.ox.ac.uk/genetics_software/snpTest/snpTest_v2.4.1.html).

In addition, we used the following R packages:

cerebroviz v1.0
EasyQC v9.2
fsbrain v0.5.5
genomicSEM v0.05
lavaan v0.6-10
LDlinkR v1.4.0
LocusZoom v2.1
Matrix v1.7-0
metafor v3.0-2
nFactors v2.4.1

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The summary statistics of social behaviour GWAS are available at <https://hdl.handle.net/1839/7a2d6cfd-431a-4e88-82f5-82ee57060958>.

1958 birth cohort. Phenotype and genetic data are provided by the UK Data Service at the University of Essex after registration (<https://doi.org/10.5255/UKDA-Series-2000032>). Information about data access is described in the CLS Data Access Framework: https://cls.ucl.ac.uk/wp-content/uploads/2017/02/CLS_Data_Access_Framework.pdf.

ABCD-NL. Due to ethical considerations, the individual data from the ABCD study cannot be deposited in a public repository. Any researcher can request the data by submitting a general publication proposal form to the ABCD Project Team as outlined at <https://www.amc.nl/web/abcd-studie-2/abcd-studie/voor-onderzoekers-.htm> under the section "Use of Data" (Dutch: Gebruik van data), by email: abcd@amsterdamumc.nl. Our data access policy can be found at <https://www.amc.nl/web/abcd-studie-2/abcd-studie/voor-onderzoekers-.htm>.

ABCD-US. The ABCD data used in this study came from the ABCD Data Release 4.0. The ABCD direct data access via NDA was discontinued in 2025 and the ABCD archive moved to the NIH Brain Development Cohorts (NBDC) Data Hub. Access to ABCD data can be obtained via the NBDC (<https://www.nbdc-datahub.org/>).

ALSPAC. The data used are available through a fully searchable data dictionary (<http://www.bristol.ac.uk/alspac/researchers/our-data/>). Access to ALSPAC data can be obtained as described within the ALSPAC data access policy (<http://www.bristol.ac.uk/alspac/researchers/access/>).

BREATHE. Access to BREATHE data can be obtained by contacting jordi.sunyer@isglobal.org or mariona.bustamante@isglobal.org.

CATSS. The data used from the CATSS cohort are not publicly available to protect sensitive phenotype information for participating children. Access to CATSS can be obtained by submitting a data request.

COPSAC. Data from the COPSAC cohort is classified as sensitive, personally identifiable data and thus cannot be made publicly available, but can potentially be made available upon request under a data transfer agreement.

DCHS. The Drakenstein Child Health Study is committed to the principle of data sharing. De-identified data will be made available to requesting researchers as appropriate. Requests for collaborations are welcome. More information can be found on our website <http://www.paediatrics.uct.ac.za/scah/dclhs>.

ELVS. The data used from the ELVS dataset are not readily available due to the ethical governance set in place for this study. The data are shared following application to investigators of the ELVS study. Requests to access the genetic datasets should be directed to Angela Morgan, angela.morgan@mcri.edu.au.

FinnTwin12. In accordance with the Finnish Biobank Act, the data used in the analysis are deposited in the Biobank of the Finnish Institute for Health and Welfare (<https://thl.fi/en/web/thl-biobank/forresearchers>). They are available to researchers from academia and companies after a written application and following the relevant Finnish legislation. To ensure the protection of privacy and compliance with national data protection legislation, a data use/transfer agreement is needed, the content and specific clauses of which will depend on the nature of the requested data.

GenR. Data from the Generation R cohort are not publicly available due to legal and ethical restrictions; however, access can be requested via the Generation R administration (secretariaat.genr@erasmusmc.nl). Requests for Generation R data access are evaluated by the Generation R Management Team.

GINILISA. Due to data protection reasons, the datasets generated and/or analysed during the current study cannot be made publicly available. The datasets are available to interested researchers from the corresponding author upon reasonable request (e.g. reproducibility), provided the release is consistent with the consent given by the GINIplus and LISA study participants. Ethical approval might be obtained for the release, and a data transfer agreement from the legal department of the Helmholtz Zentrum München must be accepted.

GLAKU. Researchers can obtain a de-identified GLAKU dataset after having obtained approval from the GLAKU Study Board.

HRS. The phenotype survey data are provided by the health and retirement study (HRS) without charge, after registration. The genomic data are provided by the health and retirement study after IRB approval and approval by NCBI-dbGaP (dbGaP Study Accession: phs000428.v2.p2), without charge. Sensitive data, including genetic information and biomarkers, are only available through an approved data application process (Health and Retirement Study, DUA Review Committee, 426 Thompson Street, Ann Arbor, Michigan 48104-2321, hrrdatareq@umich.edu). By requesting access to the HRS genetic information via NCBI-dbGaP, the authors agreed that the data are only shared within the working group permitted to work with the data (St Pourcain group) (<https://hrs.isr.umich.edu/sites/default/files/>

HRS-Genetic-Data-Access-Agreement.pdf). Information about the data access is described here: <https://hrs.isr.umich.edu/data-products/>.

INMA. INMA data (INMA-GSA and INMA-Omni1) are available by request from the Infancia y Medio Ambiente Executive Committee for researchers who meet the criteria for access to confidential data. Access to INMA data can be obtained by contacting inma@proyecto.inma.org.

INSchool. Raw data from INSchool are not publicly available because of limitations in ethical approvals, and the summary data will be available by contacting marta.ribases@vhir.org.

LSAC. Data from the LSAC cohort can be obtained by completing an online form, following instructions from the LSAC data custodian, The Australian Institute of Family Studies. Information about data access is described in <https://growingupinaustralia.gov.au/data-and-documentation/accessing-lsac-data>.

MCS. Phenotype and genetic data from the Millennium Cohort Study are provided by the UK Data Service at the University of Essex after registration (<https://doi.org/10.5255/UKDA-Series-2000031>). Information about data access is described in the CLS Data Access Framework: https://cls.ucl.ac.uk/wp-content/uploads/2017/02/CLS_Data_Access_Framework.pdf.

MoBa. Data from the Norwegian Mother, Father and Child Cohort Study and the Medical Birth Registry of Norway used in this study are managed by the national health register holders in Norway (Norwegian Institute of Public Health). The consent given by the participants does not allow for storage of data on an individual level in repositories or journals. Researchers who want access to data sets for replication should apply through helsedata.no. Access to data sets requires approval from the Regional Committee for Medical and Health Research Ethics in Norway (REC) and an agreement with MoBa. In addition, compliance with the EU General Data Protection Regulation (GDPR) is required.

MUSP. MUSP data are available from a third party upon reasonable request. Contact details can be found here: <https://social-science.uq.edu.au/mater-university-queensland-study-pregnancy?p=5#5>.

NFBC1986. NFBC data are available from the University of Oulu, Infrastructure for Population Studies. Permission to use the data can be applied for research purposes via an electronic material request portal. In the use of data, we follow the EU General Data Protection Regulation (679/2016) and the Finnish Data Protection Act. The use of personal data is based on a cohort participant's written informed consent in their latest follow-up study, which may cause limitations to its use. Please, contact the NFBC project center (NFBCprojectcenter@oulu.fi) and visit the cohort website (www.oulu.fi/nfbc) for more information.

NTR. The data used from the NTR cohort are not publicly available to protect sensitive phenotype information for participating children. Access to NTR can be obtained by submitting a data request.

The Raine Study. Restrictions apply to the availability of data generated or analysed during this study to preserve patient confidentiality. On request (<https://ross.rainestudy.org.au>), details of the restrictions and any conditions under which data access is granted may be provided.

SYS. SYS data are available upon request and should be addressed to Dr Zdenka Pausova (zdenka.pausova@sickkids.ca) and Dr Tomas Paus (tpausresearch@gmail.com). Further details about the protocol can be found at <http://www.saguenay-youth-study.org/>.

TEDS. Researchers can apply for access to the TEDS data through their data access mechanism (www.teds.ac.uk/researchers/teds-data-access-policy).

TRAILS. TRAILS data are not publicly available to researchers outside the TRAILS consortium, but can be obtained by submitting a publication proposal. For more details, please see www.trails.nl.

Additionally, we accessed the following publicly available databases: GWAS Catalog (<https://www.ebi.ac.uk/gwas/>), pheWeb (<https://pheweb.org/>), SFARI (<https://gene.sfari.org/database/human-gene/>), UK Biobank - Neale lab (<https://www.nealelab.is/uk-biobank>).

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

We use the term sex to capture genetic differences in males and females, i.e. sex represents a biological attribute inferred from genome-wide genotyping information. Sex was determined based on sex chromosome information. Biological sex was used as a covariate in our cohort-genome-wide analyses, and we report sex-combined GWAS findings, exploiting the power of our analyses.

Reporting on race, ethnicity, or other socially relevant groupings

Ancestry was inferred from genetic information to control for confounding in genetic association analyses (e.g. population stratification), and represents, as sex, a biological attribute. Ancestry does not refer to a socially relevant grouping in this work, and we refer to individuals as European-ancestry and African-ancestry individuals, following the recommendation of the Committee on the Use of Race, Ethnicity, and Ancestry (<https://www.nap.edu/catalog/26902>).

Population characteristics

We conducted a genome-wide association study (GWAS) meta-analysis combining 195 GWAS summary statistics from 73,321 individuals of European ancestry with genotype information and social behaviour, including 491,246 repeat observations (Fig. 1, Extended Data Table 1, Supplementary Table 1, Supplementary Note 1). We follow-up findings in 16,305 independent individuals with 115,254 repeat observations, of which 835 were of African ancestry and the remaining individuals of European descent. An overview of available social behaviour measures across the social domains (low prosocial behaviour and peer/social problems), as reported by different informants, spanning ages 2-29 years, is shown in Fig. 1c (see also Extended Data Table 1, Supplementary Table 1, Supplementary Note 1). On average, 50% of participants were male (Supplementary Table 1). For cohort-level univariate GWAS, the genetic association was assessed with a linear regression approach, adjusting for covariate effects (age, age squared, sex and ancestry-informative principal components). For the genome-wide meta-regression analysis, social domains (low prosocial behaviour and peer/social problems), informants and median cohort age were included as moderators.

Recruitment

Data within this study were obtained from prospective population-based or community-based (twin) cohorts within the EAGLE consortium (<https://www.eagle-consortium.org/>), which are described in detail in Supplementary Note 1.

Ethics oversight

1958BC. Oral or written consent was provided by the parents of the children participating in this study (<https://cls.ucl.ac.uk/wp-content/uploads/2017/07/NCDS-Ethical-review-and-Consent-2014.pdf>). We received ethical approval to access and analyse pre-collected de-identified genotype and phenotype data from this cohort from the Radboud University Ethics Committee Social Science (ECSW-2018-160).

ABCD-NL. All ABCD participants gave written informed consent for data collection of the phenotypes. The ABCD study protocol was approved by the Central Committee on Research Involving Human Subjects in the Netherlands, the medical ethics review committees of the participating hospitals, and the Registration Committee of the Municipality of Amsterdam.

ABCD-US. Parents or guardians provided written consent, while children provided written assent. The Institutional Review Board (IRB) at the University of California, San Diego, approved all aspects of the ABCD study. The ABCD data repository grows and changes over time. The ABCD data used in this report came from ABCD Curated Annual Release 4.0.

ALSPAC. Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. Consent for biological samples has been collected following the Human Tissue Act (2004). Informed consent for the use of data collected via questionnaires and clinics was obtained from participants following recommendations of the ALSPAC Ethics and Law Committee at the time.

BREATHE. All parents or legal guardians provided written informed consent. The study was approved by the IMIM-Parc de Salut Mar Research Ethics Committee (No. 2010/41221/I), Barcelona, Spain, and the FP7-ERC-2010-AdG Ethics Review Committee (268479-22022011).

CATSS. Informed consent was provided by participants in CATSS. The study and analyses of genetic data were approved by the Karolinska Institute Ethical Review Board and the Swedish Ethical Review Authority.

COPSAC. The participants took part in the study voluntarily and after written informed consent from the caretaker. The participating families are informed that the study is unlikely to benefit their child and that discoveries will only benefit future generations. The main study protocol for starting the cohort was approved by the Ethics Committee (H-B-2008-093) and the Danish Data Protection Agency (2008-41-2599).

DCHS. Written consent was provided by the parents of the study and is annually renewed. The study was approved by the Ethics Committee of the Faculty of Health Sciences, University of Cape Town, by Stellenbosch University and the Western Cape Provincial Research Committee.

ELVS. Written informed consent was provided by the parents, and ethical approval was obtained from the Royal Children's Hospital (23018 and 27078) and La Trobe University Human Ethics Committee (03-32).

FinnTwin12. The ethics committee of the Department of Public Health of the University of Helsinki (Helsinki, Finland), the ethics committee of the Helsinki University Central Hospital District (Helsinki, Finland, HUS/2226/2021), and the Institutional Review Board of Indiana University (Bloomington, Indiana, USA) approved the FinnTwin12 study protocol. HUS reviews the study annually, and 2023's statement is number 4/2023, dated February 1, 2023.

GenR. Participants in GenR have given written informed consent. The study protocol has been approved by the Medical Ethical Committee of the Erasmus Medical Centre, Rotterdam.

GINILISA. Written consent from participants' families and approval by the local Ethics Committees (Bavarian Board of Physicians, Medical Faculty of the University of Leipzig, Board of Physicians of North-Rhine-Westphalia) were obtained.

GLAKU. Informed consent was obtained from all participants. The study protocol was approved by the ethical committees of the City of Helsinki and the Uusimaa Hospital District.

HRS. Consent was provided by respondents before saliva collection. We received ethical approval to access and analyse pre-collected de-identified genotype and phenotype data from this cohort from the Radboud University Ethics Committee Social Science (ECSW-2018-160).

INMA_GSA. The study was approved by the Institutional Boards of the recruiting centres and by the Basque Ethics Committee.

INMA_Omni1. The study was approved by the Ethical Committee of each participating centre.

INSchool. Written informed consent was obtained from participants' parents before inclusion into the study. The study was approved by the Clinical Research Ethics Committee of the Hospital Universitari Vall d'Hebron. All methods were performed in accordance with the relevant guidelines and regulations.

LSAC. Written informed consent and assent were obtained from all parents and children. The Australian Institute of Family Studies (AIFS) Ethics Committee approved each wave of LSAC and the CheckPoint (14-26), with additional approval from Melbourne's Royal Children's Hospital Human Research Ethics Committee (33225D).

MCS. Written consent was provided by the parents of the children participating in this study (<https://cls.ucl.ac.uk/wp-content/uploads/2017/07/MCS-Ethical-review-and-consent-Shepherd-P-November-2012.pdf>). We received ethical approval to access and analyse pre-collected de-identified genotype and phenotype data from this cohort from the Radboud University Ethics Committee Social Science (ECSW-2018-160).

MoBa. Pregnant women provided written consent before enrolment in the study. The current study was approved by the Regional Committees for Medical and Health Research Ethics (2016/1702).

MUSP. At all data-collection phases of the study, mothers provided written informed consent. The current study was approved by the Mater Hospital Ethics Committee and the University of Queensland Ethics Committee.

NFBC1986. Written informed consent was provided by parents and adolescents. Studies were approved by the ethics committee of the Northern Ostrobothnia Hospital District in Oulu (EETMK: 108/2017 (15.1.2018)) and the Oulu University, Faculty of Medicine, Oulu, Finland.

NTR. Written informed consent was provided by the parents prior to enrolment in the study. The study was approved by the Central Ethics Committee on Research Involving Human Subjects of the VU University Medical Centre, Amsterdam, an Institutional Review Board certified by the U.S. Office of Human Research Protections (IRB number IRB00002991 under Federal-wide Assurance- FWA00017598; IRB/institute codes, NTR 03-180).

The Raine Study. Parents, guardians and young adult participants provided written informed consent either before enrollment or at data collection at each follow-up. Ethics approval for the original pregnancy cohort and subsequent follow-ups was granted by the Human Research Ethics Committee of King Edward Memorial Hospital, Princess Margaret Hospital, the University of Western Australia, and the Health Department of Western Australia.

SYS. Written consent of the parents and assent of the adolescents were obtained. The study protocol was approved by the Research Ethics Committee of the Chicoutimi Hospital and the Hospital for Sick Children in Toronto.

TEDS. Written informed consent was obtained from parents before data collection and from TEDS participants themselves, past the age of 18 years. King's College London's Ethics Committee approved the project for the Institute of Psychiatry, Psychology and Neuroscience, PNM/09/10–104.

TRAILS. All measurements were carried out with participants' adequate understanding and written consent. The study was approved by the Dutch Central Committee on Research Involving Human Subjects (CCMO).

Ethical approval for secondary data analyses at the Max Planck Institute for Psycholinguistics, here solely based on GWAS summary statistics, has been provided by the Ethics Committee Social Sciences at Radboud University (ECSW-2023-0107).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Within the EAGLE consortium (https://www.eagle-consortium.org/), we invited all population-based and community-based (twin) cohorts with repeated measures of social behaviour, as collected with a prospective longitudinal design, to participate in the study. No sample size for the GWAS meta-analysis was predetermined. Data from 25 independent cohorts was drawn to conduct the GWAS meta-analysis. Our meta-analysis included phenotype measures across social domains (peer/social problems and low prosocial behaviour), informants (parent, teacher, self) and ages (2 to 29 years). The total sample size for this meta-analysis included a total of 491,246 observations from 73,321 individuals of European ancestry with genotype information and data on social behaviour. The effective sample size of each SNP (Neff) was estimated from the squared ratio of the test statistic to the effect size estimate, scaled by a factor accounting for allele frequency variation, $2 \times \text{MAF} \times (1 - \text{MAF})$ (See Methods). The total effective sample size in this study is 122,239.
Data exclusions	A detailed description of quality control measures and data exclusions is provided in Supplementary Note 10, Supplementary Table 35 and the Methods section. In short, for pre-imputation quality control, individuals were excluded provided that they had a mismatch in reported and genetic sex, a missingness above 3%, an increased relatedness, or non-European genetic ancestry. Single variants were excluded if they had a missingness above 5%, deviated from Hardy-Weinberg equilibrium (p-value $< 5 \times 10^{-7}$) or were low-frequent (minor allele frequency $< 1\%$). Also, insertions/deletions, duplicate SNPs, and multi-allelic SNPs were excluded. For post-imputation quality control, variants were excluded if imputation quality scores were low, minor allele frequency was lower than 0.5%, and/or the minor allele count was ≤ 10. SNP alleles were aligned against the HRCr1.1 reference panel, and variants were excluded if not contained in the reference panel, SNP alleles were missing or mismatched, or the reported allele frequency deviated from the reported HRCr1.1 frequency. GWAS summary statistics for which heritability estimates had either an upper 95% confidence interval below zero or a lower 95% confidence interval above one were also removed from the genome-wide meta-regression analysis to exclude unreliable measures.
Replication	We conducted a proof-of-concept replication of the meta-analysed social behaviour GWAS by computing polygenic scores (PGS) in three independent cohorts (Millennium Cohort Study, 1958 birth cohort and Drakenstein Child Health Study) with data on social behaviour. We included 16,305 independent individuals with 115,254 repeat observations of either European or African ancestry and carried out extensive validation analyses of polygenic scores (PGS), especially in the Millennium Cohort Study, to confirm the polygenic predictability and accuracy.

The predictive accuracy of PGS was high, predicting variation in social domains, informants and, largely, in developmental stages, except for the predictive accuracy of self-reported low prosocial behaviour. Here, the outcome was better captured by parent-report PGS, not the matching self-report PGS (Supplementary Table 17, Supplementary Figure 16). Self-reported prosocial behaviour measures are prone to socially desirable responding, and this bias may affect genomic analyses. Given the meta-regression character of the study, reported GWAS signals are based on all available samples, capturing variation across meta-regression moderators. Follow-up analyses of GWAS signals would, therefore, require the integration of the follow-up cohorts into the meta-regression model, invalidating the independence of the follow-up cohorts for PGS validation analyses, and were, consequently, not conducted.

Randomization Randomisation was not conducted due to the prospective longitudinal design of the studied cohorts and the adopted GWAS approach.

Blinding Blinding was not conducted due to the prospective longitudinal design of the studied cohorts and the adopted GWAS approach.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging