

An axis of genetic heterogeneity in autism is indexed by age at diagnosis and is associated with varying developmental and mental health profiles

Xinhe Zhang^{1,2}, Jakob Grove³⁻⁶, Yuanjun Gu^{1,2}, Cornelia K. Buus⁵, Lea K. Nielsen⁵, Sharon A.S. Neufeld¹, Mahmoud Koko⁷, Daniel S Malawsky⁵, Emma Wade⁵, Ellen Verhoef⁸, Anna Gui^{9,10}, Laura Hegemann¹¹⁻¹³, APEX consortium, iPSYCH Autism Consortium, PGC-PTSD Consortium, Daniel H. Geschwind¹⁴⁻¹⁷, Naomi R. Wray^{18,19}, Alexandra Havdahl¹¹⁻¹³, Angelica Ronald^{10,20}, Beate St. Pourcain^{8,21,22}, Elise B. Robinson^{23,24}, Thomas Bourgeron²⁵, Simon Baron-Cohen^{1,2,26}, Anders D. Børglum³⁻⁵, Hilary C. Martin⁷, Varun Warriar^{1,2,26}

1. Department of Psychiatry, University of Cambridge
2. Autism Research Centre, Department of Psychiatry, University of Cambridge, Cambridge, UK
3. The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Aarhus, Denmark
4. Center for Genomics and Personalized Medicine (CGPM), Aarhus University, Aarhus, Denmark
5. Department of Biomedicine (Human Genetics) and iSEQ Center, Aarhus University, Aarhus, Denmark
6. Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark
7. Human Genetics Programme, Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, CB10 1SA, UK
8. Language and Genetics Department, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
9. Department of Psychology, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, United Kingdom
10. Centre for Brain and Cognitive Development, Department of Psychological Sciences, Birkbeck University of London, London, WC1E 7HX, United Kingdom
11. Department of Psychology, University of Oslo, Oslo, Norway
12. Nic Waals Institute, Lovisenberg Diaconal Hospital, Oslo, Norway
13. PsychGen Center for Genetic Epidemiology and Mental Health, Norwegian Institute of Public Health, Oslo, Norway
14. Program in Neurobehavioral Genetics and Center for Autism Research and Treatment, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA.
15. Program in Neurogenetics, Department of Neurology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA.
16. Department of Psychiatry, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA.
17. Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA.
18. Institute for Molecular Bioscience, University of Queensland, Brisbane, QLD, Australia
19. Department of Psychiatry, University of Oxford, Oxford, UK
20. School of Psychology, University of Surrey, Guildford, Surrey, GU2 7XH, United Kingdom

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

- 1 21. MRC Integrative Epidemiology Unit, University of Bristol, United Kingdom.
- 2 22. Donders Institute for Brain, Cognition and Behaviour, Radboud University,
- 3 The Netherlands.
- 4 23. Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA,
- 5 USA
- 6 24. The Broad Institute of MIT and Harvard, Cambridge, MA, USA
- 7 25. Human Genetics and Cognitive Functions, Institut Pasteur, UMR3571 CNRS,
- 8 IUF, Université Paris Cité, Paris, France
- 9 26. Department of Psychology, University of Cambridge, Cambridge, UK

10
11 Correspondence to Xinhe Zhang (xz452@cam.ac.uk) or Varun Warriar
12 (vw260@cam.ac.uk)

13
14

1 There is growing recognition that earliest signs of autism need not clearly manifest in
2 the first three years of life. To what extent is this variation in developmental trajectories
3 associated with age at autism diagnosis? Does the genetic profile of autism vary with
4 age at autism diagnosis? Using longitudinal data from four birth cohorts, we
5 demonstrate that two different trajectories of socio-emotional behaviours are
6 associated with age at diagnosis. We further demonstrate that the age at autism
7 diagnosis is partly heritable ($h^2_{\text{SNP}} = 0.12$, s.e.m = 0.01), and is associated with two
8 moderately correlated ($r_g = 0.38$, s.e.m = 0.07) autism polygenic factors. One of these
9 factors is associated with earlier diagnosis of autism, lower social and communication
10 abilities in early childhood. The second factor is associated with later autism diagnosis,
11 increased socio-emotional difficulties in adolescence, and has moderate to high
12 positive genetic correlations with Attention-Deficit/Hyperactivity Disorder, mental
13 health conditions, and trauma. Overall, our research identifies an axis of heterogeneity
14 in autism, indexed by age at diagnosis, which partly explains heterogeneity in autism
15 and the profiles of co-occurring neurodevelopmental and mental health profiles. Our
16 findings have important implications for how we conceptualise autism and provide one
17 model to explain some of the diversity within autism.

18
19

20 **Main**

21 Autism is a term used to describe a group of conditions characterised by difficulties in
22 social-communication, unusually restricted and repetitive interests, and sensory
23 differences¹. Ever since its earliest descriptions in the 1940s^{2,3}, autism has been
24 thought of as a condition that typically emerges and is diagnosed in early childhood.
25 However, recent studies demonstrate that more autistic individuals are now receiving
26 an autism diagnosis from mid-childhood onwards than in early childhood⁴⁻⁶.

27

28 One factor that may explain these findings is a shift in the conceptualisation of autism
29 over time. There is a growing recognition that the signs of autism may not clearly
30 manifest in the first three years of life^{1,7-9}, which has been recognised by the changes
31 to the diagnostic criteria for autism by DSM-5 and ICD-11. Supporting this, several
32 studies have demonstrated that a subset of children who do not initially meet the
33 criteria for an autism diagnosis later receive a diagnosis^{7,10-14}

34

35 These findings pose a series of fundamental questions regarding the aetiology of
36 autism. For instance, given the substantial heritability of developmental phenotypes¹⁵⁻
37 ¹⁷, to what extent does the genetic profile of autism vary with age at receiving an autism
38 diagnosis? How does the developmental variation in the emergence of autism
39 features contribute to age at autism diagnosis, and consequently, the genetic
40 heterogeneity within autism? Is the higher prevalence of mental health diagnosis
41 among autistic individuals diagnosed later in life^{18,19} partly due to genetic factors?

42

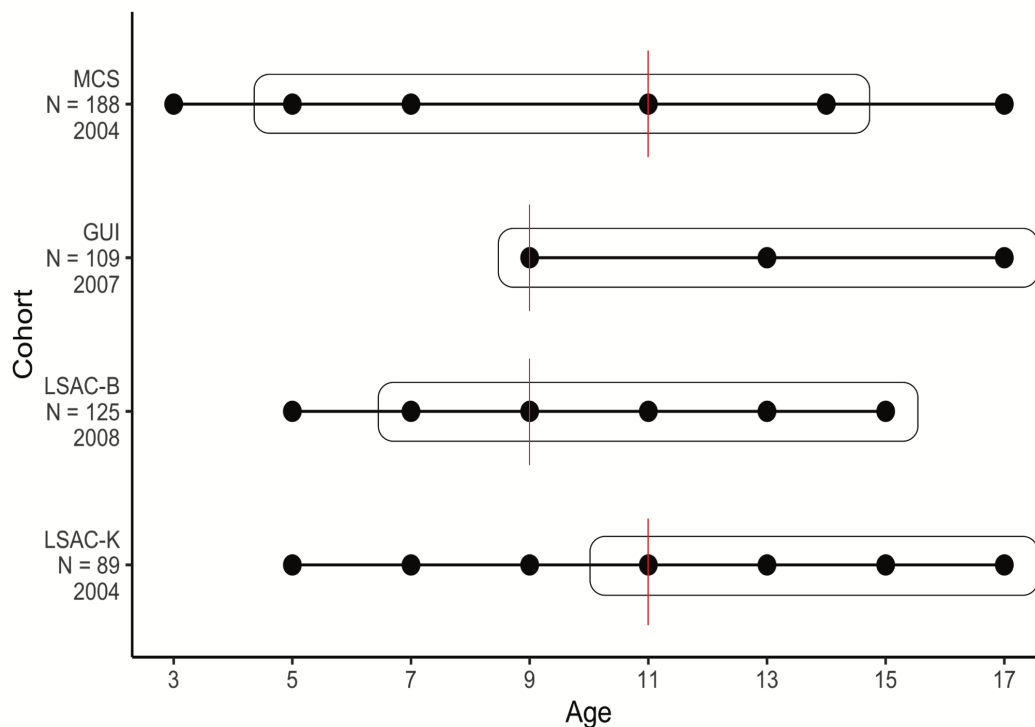
43 We address these questions using multiple epidemiological and genetic datasets.
44 Using longitudinal data from birth cohorts, we demonstrate that individual differences
45 in socio-behavioural trajectories are associated with age at autism diagnosis. We
46 further demonstrate that age at autism diagnosis is heritable, and this heritability can
47 be partly explained by two correlated polygenic autism factors. The two polygenic
48 factors are differentially associated with developmental and mental health profiles, and
49 partly explain the genetic heterogeneity of autism. We provide a summary of the study

1 and address potential questions regarding the implications of the findings in the
2 **Supplementary FAQs**.

3 4 5 **Socio-behavioural trajectories are linked to varying age at autism** 6 **diagnosis**

7 We first investigated the association between variable developmental trajectories and
8 age at autism diagnosis using four birth cohorts. This included Growing Up in Ireland
9 (GUI, born in 1998), Millennium Cohort Study (MCS, 2000), and Longitudinal Study of
10 Australian Children: Kindergarten cohort (LSAC-K, 1999) and Birth cohort (LSAC-B,
11 2003) (**Supplementary Table 1, Extended Figure 1, Supplementary Note 1**). All
12 four cohorts collected longitudinal information on socio-behaviour measured using the
13 caregiver-reported Strengths and Difficulties Questionnaire (SDQ)²⁰ and its subscales,
14 and autism diagnosis in data collection sweeps (hereafter “sweeps”) at different ages.
15 The SDQ is widely used, has excellent psychometric properties^{21–23}, and is largely
16 invariant across age, sex, and different populations^{24–26}, meaning that it is measuring
17 the same latent trait across these demographic variables.

18
19 Given increasing number of autistic individuals being diagnosed in adolescence^{4,5}, we
20 wondered if there are broad differences in the trajectories of SDQ total and subscale
21 scores among autistic individuals diagnosed before the ages of 9 - 11 (childhood
22 diagnosed group, N = 39 - 118 across cohorts, **Supplementary Table 1**) and after
23 (adolescent diagnosed group, N = 27 -73 across cohorts) (**Methods**). This age cutoff
24 period corresponds to the onset of puberty, the transition from primary to secondary
25 school, and the beginning of an increase of incidence in diagnosis of autism in
26 girls^{27,28,29}. The specific cutoff age was cohort-dependent, as different birth cohorts
27 collected information on autism diagnosis at different ages.
28



29
30 **Extended Figure 1:** Schematic diagram of the cohorts included in the study and the ages when data
31 was collected for SDQ scores (dots) and autism diagnosis (in boxes). MCS = Millennium Cohort Study;

1 GUI = Growing up in Ireland (cohort '98); LSAC-B = Longitudinal Study of Australian Children (Birth
2 cohort); LSAC-K = Longitudinal Study of Australian Children (Kindergarten cohort). Sample sizes and
3 the year of initial SDQ data collection for each cohort are shown on the ordinate axis. The age cutoff
4 used in the Latent Growth Curve Models for each cohort is indicated by a red line.
5

6 We used Latent Growth Curve Models to linearly model the trajectories of SDQ total
7 score and subscales in all four cohorts for both childhood and adolescent diagnosed
8 groups. Across four cohorts, Latent Growth Curve Models identified different
9 trajectories of SDQ total scores between the childhood and adolescent diagnosed
10 groups (Mean scores in **Figure 1A-C, Supplementary Figures 1 - 6, Supplementary**
11 **Table 2**). Similar results were obtained for peer relationship problems and prosocial
12 behaviours in all four cohorts (**Supplementary Figures 4 and 5**). Compared with
13 individuals without an autism diagnosis at any time point, the childhood diagnosed
14 autistic group had higher difficulties in early childhood that remained relatively stable
15 or gently declined in adolescence. Compared to the childhood diagnosed autistic
16 group, the adolescent diagnosed autistic group had fewer difficulties during early
17 childhood, but difficulties increased in later childhood and persisted into adolescence.
18

19 In MCS, we ran a series of sensitivity analyses to check the robustness of the above
20 results. We obtained consistent results: (1) in an expanded sample of MCS which
21 included autistic individuals with co-occurring Attention-Deficit/Hyperactivity Disorder
22 (ADHD) and inconsistent reports of an autism diagnosis (MCS-expanded, **Methods,**
23 **Supplementary Table 2**); (2) after imputing missing SDQ scores (MCS-imputed)
24 (**Supplementary Table 3, Supplementary Note 2**); and (3) when restricting to males.
25 In all four birth cohorts, models stratified by age at diagnosis were generally a better
26 fit to the data than sex-stratified models (**Supplementary Table 2**), suggesting that the
27 results do not primarily reflect sex differences in age at autism diagnosis
28 (**Supplementary Note 3**).
29

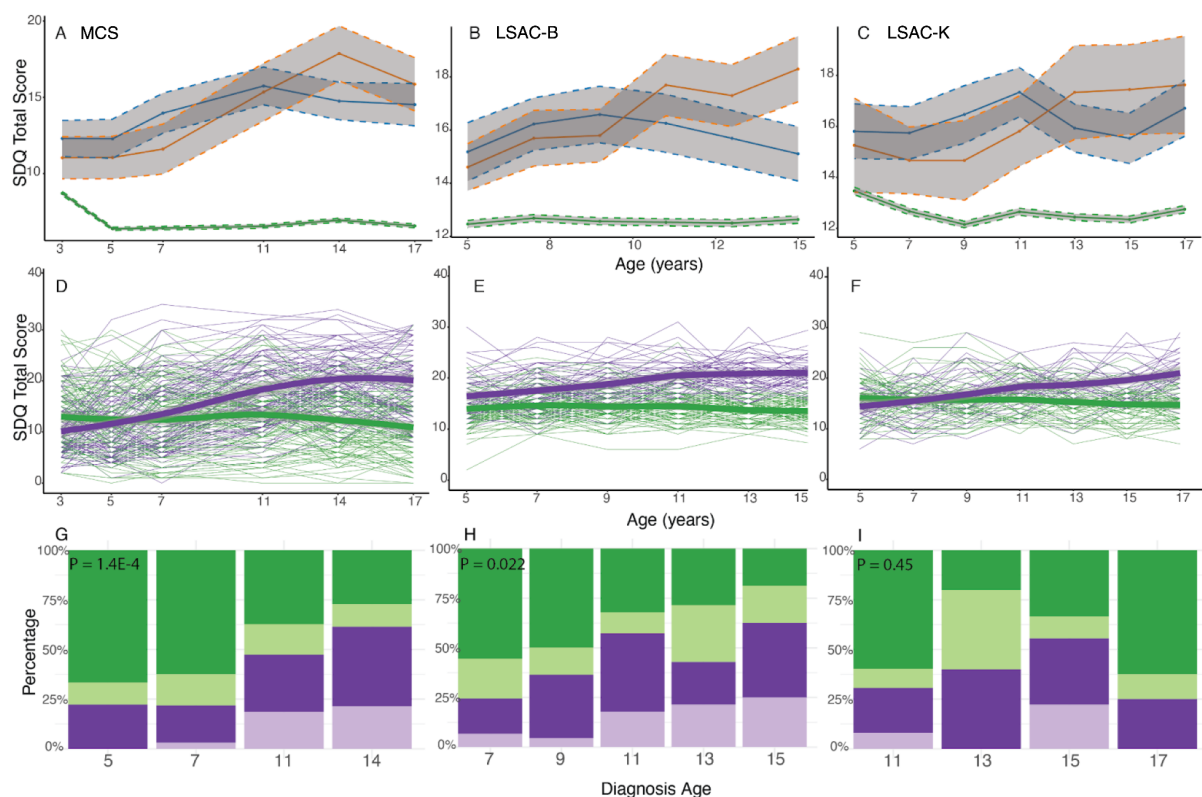
30 To assess the specificity for autism, we ran Latent Growth Curve Models on SDQ total
31 scores and subscales with children with an ADHD but without an autism diagnosis in
32 the MCS cohort (N = 89). Children with ADHD diagnosed in childhood and
33 adolescence differed only nominally in the slopes of the hyperactivity/inattention (P =
34 0.026) and prosocial behaviour subscales (P = 0.029) (**Supplementary Table 4 and**
35 **Supplementary Figures 7 and 8**). In MCS, compared to adolescent diagnosed
36 children with ADHD, adolescent diagnosed autistic children had a steeper increase in
37 peer relationship problems (P = 5.77×10^{-3}) and emotional symptoms (P = 0.012)
38 across development. However, these results must be interpreted cautiously given the
39 low number of children with only ADHD in MCS.
40

41 Recognising that the age at diagnosis threshold used to categorise autistic individuals
42 into two groups is to some extent arbitrary, we used Growth Mixture Models (GMMs)
43 to identify latent trajectories of SDQ total and subscale scores among autistic
44 individuals in all four cohorts. GMMs do not require *a priori* grouping but identify
45 subgroups based on longitudinal changes in SDQ scores .
46

47 Across three of the four birth cohorts, GMMs identified two-trajectory models as being
48 optimal for SDQ total scores and a majority of subscale scores (**Supplementary Table**
49 **5**). The exception was the GUI cohort, where a one-trajectory model was optimal, likely
50 due to fewer sweeps (three) for SDQ scores. Amongst the other cohorts, one latent
51 trajectory was characterised by difficulties in early childhood, which remained stable

1 or gently declined with age (early childhood emergent latent class). The other latent
 2 trajectory was characterised by fewer difficulties in early childhood which increased in
 3 late childhood and adolescence (late childhood emergent latent class) (**Figure 1 D-F**).
 4 Autistic individuals in the early childhood emergent latent class were more likely to be
 5 diagnosed in childhood compared to autistic individuals in the late childhood emergent
 6 latent class in MCS ($P = 1.43 \times 10^{-4}$, chi-square test) and LSAC-B ($P = 0.022$, chi-square
 7 test) (**Figure 1 G-I, Supplementary Table 6**), but this difference was not significant in
 8 LSAC-K, possibly because age 11 was the earliest when an autism diagnosis was
 9 recorded.

10
 11 Similar results were obtained for all the SDQ subscales in MCS, LSAC-B and LSAC-
 12 K except for conduct and peer relationship problems in LSAC-K, where no distinct
 13 trajectory groups were identified (**Supplementary Table 5, Supplementary Figures**
 14 **1 - 5**). In MCS, these results were largely consistent: (1) in the expanded sample after
 15 including individuals with co-occurring ADHD and inconsistent autism diagnoses
 16 (**Supplementary Table 4**); (2) after imputation (**Supplementary Table 3,**
 17 **Supplementary Note 4**); and (3) when restricting to only males (**Supplementary**
 18 **Table 5**). In contrast, although we obtained two latent classes based on SDQ and
 19 subscale trajectories among individuals with a diagnosis of ADHD but not autism,
 20 these latent classes were not significantly associated with age at ADHD diagnosis
 21 (**Supplementary Table 4, Supplementary Figures 7 and 8**).



22
 23 **Figure 1: Trajectory analyses in three of the four birth cohorts.** A - C: Mean SDQ total scores in
 24 autistic individuals diagnosed in childhood (blue) and adolescence (orange), and individuals without an
 25 autism diagnosis (green) in the MCS (A), LSAC-B (B), and LSAC-K (C) cohorts. Grey regions indicate
 26 95% confidence intervals. D - F: Longitudinal growth mixture models of SDQ total scores among autistic
 27 individuals, demonstrating the presence of two groups (green indicating early childhood emergent latent
 28 class and purple indicating late childhood emergent latent class) in the MCS (D), LSAC-B (E) and LSAC-

1 K (F) cohorts. G - I: Stacked bar charts providing the proportion of individuals who had been diagnosed
2 as autistic at specific ages, categorised by membership in the latent classes identified from the growth
3 mixture models in MCS (G), LSAC-B (H), and LSAC-K (I) cohorts. Darker colours indicate males and
4 lighter colours indicate females. P-values (inset) are from chi-square tests comparing the distribution of
5 age at autism diagnosis between the two latent classes. Results from GUI have not been plotted here
6 and are available in **Supplementary Figure 6**.

7
8 In all three cohorts, sex ratio in the late childhood emergent latent class of SDQ total
9 scores compared to the early childhood emergent class was statistically similar (male:
10 female ratio = 3.92 - 2.11; $P > 0.05$, chi-square tests, **Supplementary Table 6**). After
11 accounting for sex, individuals in the late childhood emergent latent class were more
12 likely than those in the early childhood emergent class to have higher depressive
13 symptoms measured using the Short Mood and Feelings Questionnaire³⁰ (MCS-C, P
14 = 3.84×10^{-4} ; LSAC-B, $P = 9.82 \times 10^{-13}$; LSAC-K, $P = 1.86 \times 10^{-6}$), have higher rates of
15 diagnosed anxiety (LSAC-B, $P = 5.54 \times 10^{-7}$; LSAC-K, $P = 3.99 \times 10^{-3}$) and depression
16 (LSAC-B, $P = 3.67 \times 10^{-4}$; LSAC-K, $P = 4.69 \times 10^{-3}$), were more likely to self-harm (LSAC-
17 B, $P = 0.018$), or have higher rates of suicidal ideation (MCS, $P = 2.50 \times 10^{-3}$)
18 (**Supplementary Table 7**).

19
20 Given the significant association between age at autism diagnosis and the GMM latent
21 class membership in LSAC-B and MCS, we wondered if these socio-behavioural
22 trajectories explain any variance in age at autism diagnosis after accounting for socio-
23 demographic factors (e.g., sex, ethnic minority, SES, living area deprivation) and
24 child's cognitive aptitude, all associated with age at autism diagnosis^{31,32}. Multiple
25 linear regression models indicated that latent class membership from the GMM, socio-
26 demographic factors, and child's cognitive aptitude together accounted for 17.4%
27 (LSAC-B) - 35.0% (MCS) of the variance in age at autism diagnosis (**Supplementary**
28 **Table 8**). Latent classes of SDQ total and subscale scores alone explained 9.9%
29 (LSAC-B) - 30.3% (MCS) of total variance. In MCS-expanded, the full model explained
30 14.8% of the variance, and SDQ total and subscale scores latent class memberships
31 explained 10.0% of the variance. In MCS-imputed, the full model explained 59.8% of
32 the variance and the latent class memberships of SDQ total and subscale scores
33 explained 56.6% of the variance (**Supplementary Table 3**).

34
35 We further investigated if the effects of socio-demographic factors and cognitive
36 aptitude on age at diagnosis were partly mediated by SDQ latent classes, but did not
37 identify a significant mediated effect. This suggests that the effects of the demographic
38 variables on age at autism diagnosis is largely independent of the effects of SDQ latent
39 classes (**Supplementary Table 8, Supplementary Note 5**).

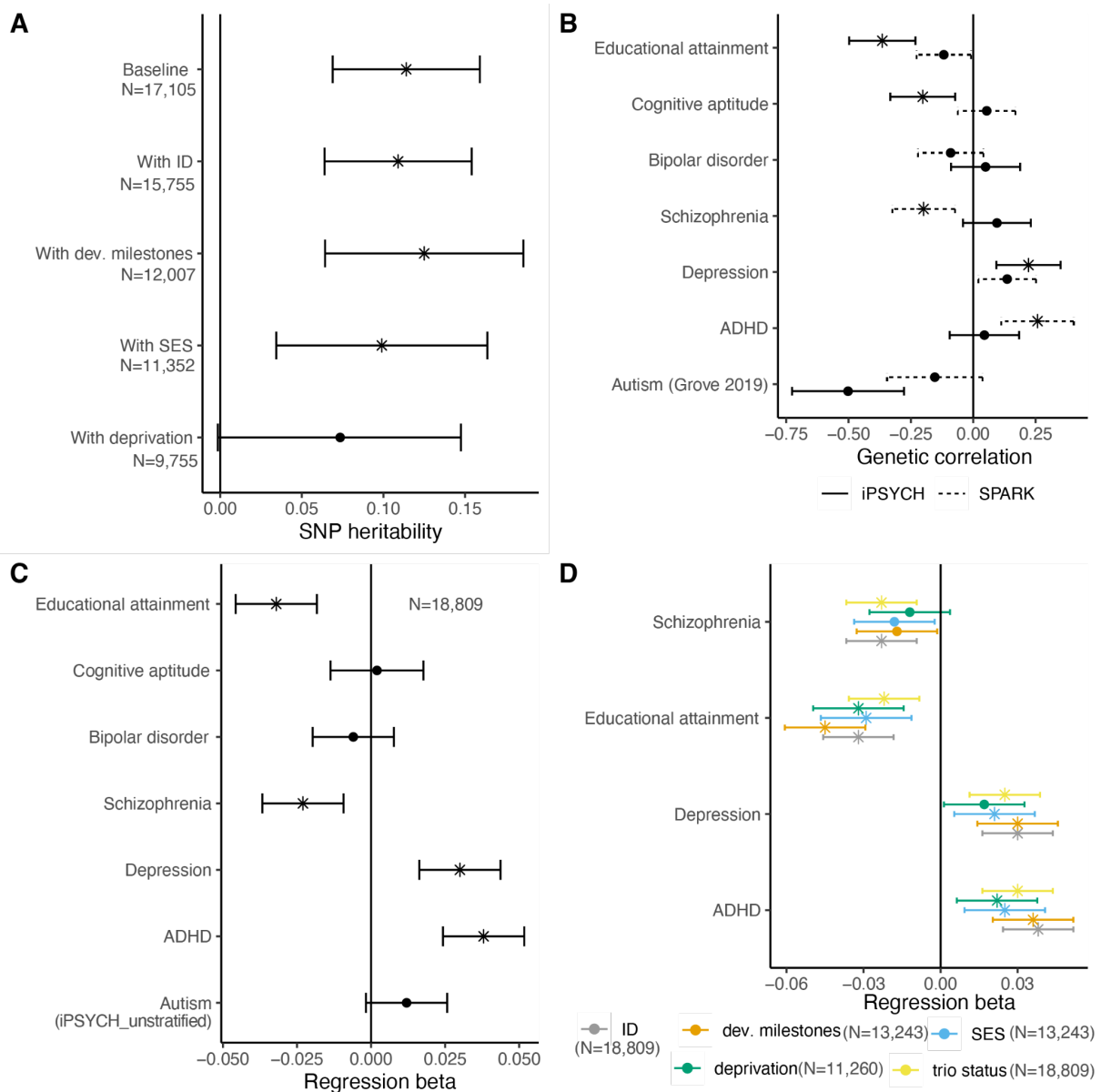
42 **Age at autism diagnosis is partly genetic**

43 The above analyses demonstrate that variation in socio-behavioural trajectories,
44 measured using the SDQ, is associated with age at autism diagnosis. Previous
45 research has demonstrated that developmental variation in socio-behavioural profiles
46 is partly explained by genetic factors³³⁻³⁸. A corollary of this is that genetic factors may
47 also be associated with age at autism diagnosis.

48
49 We tested this in a US-based cohort of autistic individuals (SPARK³⁹ : $N_{\max} = 17,105$)
50 where we identified a significant heritability estimated from single nucleotide
51 polymorphisms (SNP-based heritability) for age at autism diagnosis ($h^2_{\text{SNP}} = 0.11$,

1 s.e.m = 0.01). This heritability did not significantly attenuate after accounting for the
 2 child's developmental phenotypes (age of walking, age at first words, and intellectual
 3 disability [ID]), parental socio-economic status that may correlate with greater parental
 4 awareness and ability to access diagnostic services, and neighbourhood
 5 socioeconomic deprivation which may index availability of healthcare services (**Figure**
 6 **2A, Supplementary Table 9**).
 7

8 We used genetic correlation and polygenic score (PGS) analyses to better
 9 characterise the genetics of age at autism diagnosis ($N_{\max} = 18,809$). Later age at
 10 autism diagnosis was significantly positively genetically correlated with ADHD⁴⁰, and
 11 negatively with schizophrenia⁴¹ (**Figure 2B and 2C, Supplementary Table 10**). In
 12 SPARK, PGS for ADHD and educational attainment remained significantly associated
 13 with age at diagnosis after accounting for ID, developmental milestones, socio-
 14 economic status and neighbourhood deprivation, and trio status (i.e, two parents and
 15 one child, **Figure 2D, Supplementary Table 11**).
 16
 17



18
 19

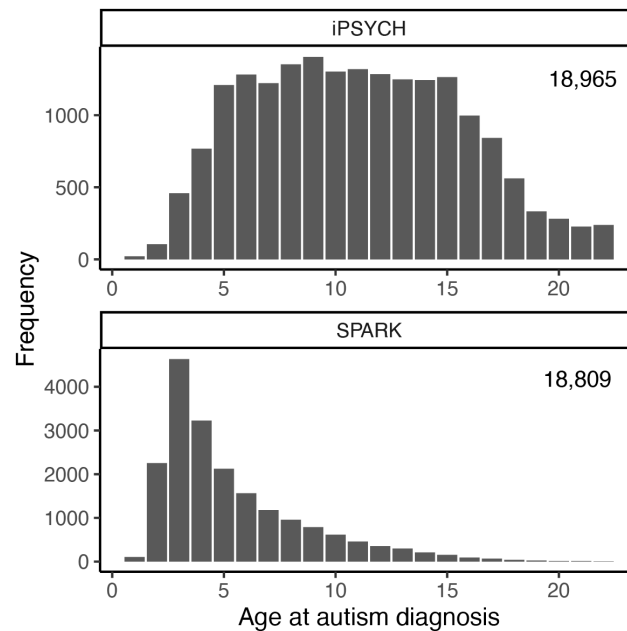
1 **Figure 2: Genetic correlates of age at autism diagnosis** A. SNP-based heritability of age at autism
2 diagnosis with and without the inclusion of covariates. B. Genetic correlation between age at autism
3 diagnosis genome-wide association studies (GWAS) from SPARK and iPSYCH and psychiatric,
4 neurodevelopmental, and cognitive phenotypes. The Autism GWAS used is the Grove et al., 2019
5 GWAS. C. Association between PGS for selected neurodevelopmental, cognitive and psychiatric
6 phenotypes and age at autism diagnosis in SPARK. Estimates provided after correcting for ID, sex, age
7 at recruitment into the study, and 10 genetic principal components. Sample sizes provided in inset. D.
8 For those significant in C, association of PGS for schizophrenia, educational attainment, depression,
9 cognitive aptitude and ADHD with age at autism diagnosis in baseline models and after correcting for
10 intellectual disability (ID), developmental (dev.) milestones, socio-economic status (SES) and
11 deprivation. Sample sizes provided in parenthesis. For all plots, points indicate the estimate, whiskers
12 indicate 95% confidence intervals. For plots B-D points with an asterisk (*) indicate significant
13 associations with Benjamini-Yekutieli (BY) adjustment. For plot A, an asterisk (*) indicates significance
14 at $P < 0.05$ as no multiple testing adjustment is needed for the sensitivity analyses, and whiskers
15 indicate 95% confidence intervals.
16

17 As females are significantly more likely to be diagnosed as autistic later than males²⁷,
18 we investigated if there is an interaction effect between sex and the PGS for four
19 phenotypes significantly associated with age at autism diagnosis. We did not identify
20 any significant PGS-by-sex interaction effects (**Supplementary Table 12**).
21

22 Parents play an important role in recognising autistic features in their children and
23 pursuing an autism diagnosis. Consequently, we wondered if the associations
24 between ADHD and educational attainment (EA) PGS on age at autism diagnosis
25 were due to parental indirect genetic effects (where parental genetics impacts child's
26 age at autism diagnosis via parental behaviours) or child's direct genetic effects. In
27 6,554 trios we observed significant direct effects of ADHD ($P = 0.015$) and significant
28 indirect effects of EA PGS ($P = 6.54 \times 10^{-4}$) on increasing age at autism diagnosis
29 respectively (**Supplementary Table 13**). These indirect effects must be interpreted
30 cautiously as they are not immune to confounding, including participation bias.
31

32 Previous research has demonstrated that autistic individuals are enriched for *de novo*
33 and rare inherited protein truncating or missense variants in genes intolerant to loss
34 of function mutations (constrained genes)^{42,43}. Using trios ($N = 6,206$), we
35 investigated if rare *de novo* or inherited protein truncating variants or missense
36 variants in highly constrained genes were associated with age at autism diagnosis.
37 We found no significant associations with age at autism diagnosis for either type of
38 variants (**Supplementary Table 11**). This may possibly reflect later autism diagnosis
39 in some carriers of *de novo* mutations due to diagnostic overshadowing by co-
40 occurring ID or global developmental delay.
41

42 We examined the generalisability of our findings in the Danish iPSYCH cohort (18,965
43 autistic individuals, mean diagnosis age 10.98 years, $SD = 4.64$; compared to
44 SPARK's 4.97 years, $SD = 3.28$, **Extended Figure 2**). A GWAS of age at autism
45 diagnosis in iPSYCH ($N = 19,931$) showed similar LDSC-based SNP heritability (h^2_{SNP}
46 $= 0.10$, $s.e.m = 0.03$) to SPARK ($h^2_{SNP} = 0.09$, $s.e.m = 0.03$) and moderate genetic
47 correlation between cohorts ($r_g = 0.51$, $s.e.m = 0.19$, $P = 7.56 \times 10^{-3}$). The iPSYCH
48 GWAS had positive genetic correlation with depression⁴⁴ and negative correlation
49 with educational attainment⁴⁵ and cognitive aptitude⁴⁶. Differences in genetic
50 correlations may be due to varying age distributions and potential participation bias in
51 SPARK.
52



1
2 **Extended Figure 2: Distribution of age at autism diagnosis in SPARK and iPSYCH.** Frequency
3 histograms of age at autism diagnosis in iPSYCH and SPARK. Sample sizes have been provided in
4 the inset.
5
6

7 **Characterising the genetic relationship between age at autism** 8 **diagnosis and autism**

9 The previous findings collectively demonstrate that age at autism diagnosis is heritable
10 but with complex genetic correlations that vary by cohort. Subsequently, we sought to
11 characterise the genetic relationship between age at autism diagnosis and autism.
12

13 Across phenotypes such as schizophrenia⁴⁷ and depression⁴⁸, the age at
14 diagnosis/onset is largely negatively genetically correlated with the phenotype itself⁴⁹.
15 This indicates that earlier diagnosis/onset is associated with greater polygenic
16 propensity for the condition. However, with autism, we observed variable genetic
17 correlations between age at autism diagnosis and different GWAS of autism, including
18 a nominally significant positive genetic correlation with the females-only iPSYCH⁵⁰
19 autism GWAS and the SPARK age at autism diagnosis GWAS (**Figure 3A,**
20 **Supplementary Table 14**). In addition, both age of diagnosis GWAS had moderate
21 negative genetic correlations with both the Psychiatric Genomics Consortium 2017
22 (PGC-2017) autism GWAS⁵¹ and a GWAS of autism in SPARK⁵², consistent with the
23 average age at autism diagnosis in PGC-2017 and SPARK being lower than that of
24 iPSYCH. The pattern of genetic correlations between age at autism diagnosis and the
25 various autism GWAS does not align well with differences in sex-ratio among the
26 GWAS (**Supplementary Figure 9**), but does align reasonably well with the median
27 age at diagnosis for the autism GWAS.
28

29 Given these varying genetic correlations, we wondered if the polygenic signal for age
30 at autism diagnosis reflects a mixture of different age-dependent polygenic traits.

31 To test this, we conducted GWAS of autism within the iPSYCH dataset, stratifying
32 participants into two groups: those diagnosed before age 11 (iPSYCH_{before11}, $N_{\text{autistic}} =$
33 9,500) and those diagnosed at age 11 or later (iPSYCH_{after10}, $N_{\text{autistic}} =$ 9,231). This
34 roughly coincided with the age window of 9 - 11 where we find an increase in SDQ

1 difficulties in the adolescent diagnosed group. We identified moderate positive genetic
2 correlation ($r_g = 0.70$, s.e.m = 0.06) between the two GWAS, which was significantly
3 less than 1 ($P = 3.68 \times 10^{-7}$). To provide further resolution based on age at diagnosis,
4 we also generated two additional, smaller GWAS of autism stratified by age at
5 diagnosis in iPSYCH: autism diagnosed before age nine (iPSYCH_{before9}, $N_{\text{autistic}} =$
6 5,451) and autism diagnosed after age 11 (iPSYCH_{after11}, $N_{\text{autistic}} = 8,260$). We used
7 the same group of unrelated individuals without an autism diagnosis as controls for all
8 four GWAS ($N_{\text{control}} = 36,667$).

9
10 Both genetic correlation and PGS association identified positive shared genetics
11 between iPSYCH_{after10} autism and age at autism diagnosis (**Figure 3A and B,**
12 **Supplementary Tables 14 and 15**), confirming the validity of the age of diagnosis
13 GWAS. Further sensitivity analyses confirmed that the PGS association could not be
14 explained by: (1) Diagnostic overshadowing due to ADHD; (2) Trio status; or (3)
15 Changes in diagnostic criteria between DSM-IV and DSM-5.

16
17 Furthermore, in SPARK, both sets of PGS were over-transmitted from parents to their
18 autistic children with a larger over-transmission of iPSYCH_{before11} compared to
19 iPSYCH_{after10} PGS ($P = 1.28 \times 10^{-3}$) (**Supplementary Table 15, Figure 3C**). This is
20 consistent with 90% of autistic individuals in SPARK being diagnosed before age 10.
21 We observed consistent results after stratifying by sex and ID (**Supplementary Table**
22 **15**). Finally, in MCS, iPSYCH_{before11} but not iPSYCH_{after10} was associated with autism
23 diagnosed before age 11 (**Supplementary Table 16, Figure 3D**). Taken together, this
24 suggests that although both sets of PGS are associated with autism, their effects on
25 autism vary by age at diagnosis.
26

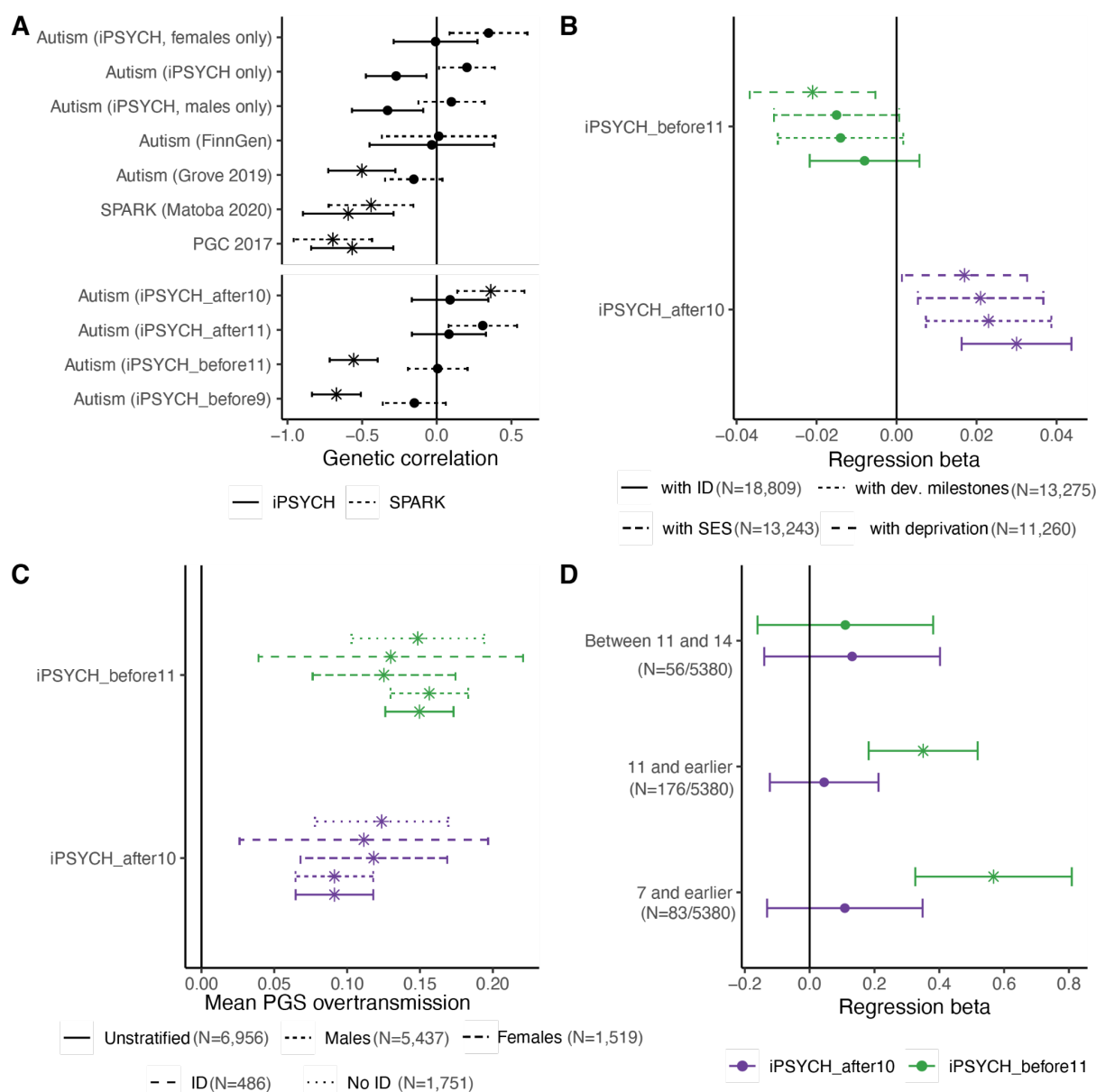
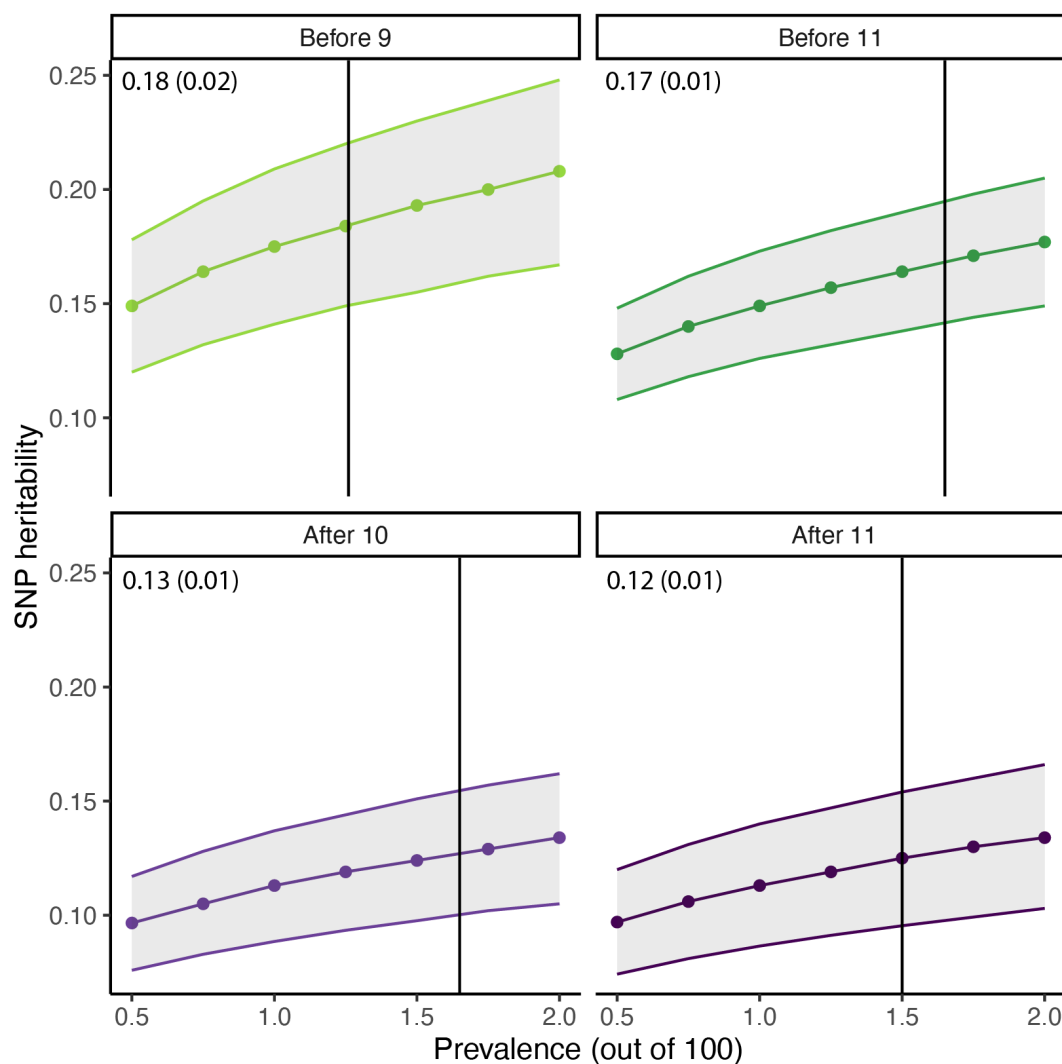


Figure 3: Genetic correlates of age at diagnosis stratified autism GWAS. A. Genetic correlation between age at autism diagnosis in SPARK and different autism GWAS. Sample sizes are provided in **Supplementary Table 14**. B. Association between age at autism diagnosis PGS for iPSYCH_{before11} and iPSYCH_{after10} in the SPARK cohort. Estimates provided after correcting for ID, developmental (dev.) milestones, socio-economic status (SES) and deprivation. C. Over-transmission of PGS for iPSYCH_{before11} and iPSYCH_{after10} from parents to autistic children in the SPARK cohort. Estimates provided for unstratified and sex-stratified analyses. Children's PGS have been standardised to parental mean PGS, with the line at zero indicating no over-transmission. D. Association between autism diagnosed in childhood and adolescence and PGS for iPSYCH_{before11} and iPSYCH_{after10} GWAS in the MCS cohort. For all plots, points indicate the estimate, whiskers indicate 95% confidence intervals. For graphs A, C, and D, points with an asterisk (*) indicate significant associations with Benjamini-Yekutieli adjustment. N indicates sample size. For graph B, points with an asterisk (*) indicate significant association after Bonferroni correction within each sensitivity analysis. N indicates sample size. For graph D, sample sizes are provided as N_{autistic}/N_{nonautistic}.

Across a range of prevalence estimates, iPSYCH_{before11} had moderately higher SNP-based heritability ($h^2 = 0.18$, s.e.m = 0.02) compared to the iPSYCH_{after10} GWAS ($h^2 = 0.13$, s.e.m = 0.01) (**Supplementary Table 17, Extended Figure 3**). The heritability of iPSYCH_{before11} was statistically similar to SPARK ($h^2 = 0.19$, s.e.m = 0.03) and PGC-

1 2017 ($h^2 = 0.20$, s.e.m = 0.02), suggesting that stratifying by age at diagnosis identifies
 2 similar SNP-based heritability between iPSYCH, SPARK, and PGC-2017.
 3
 4

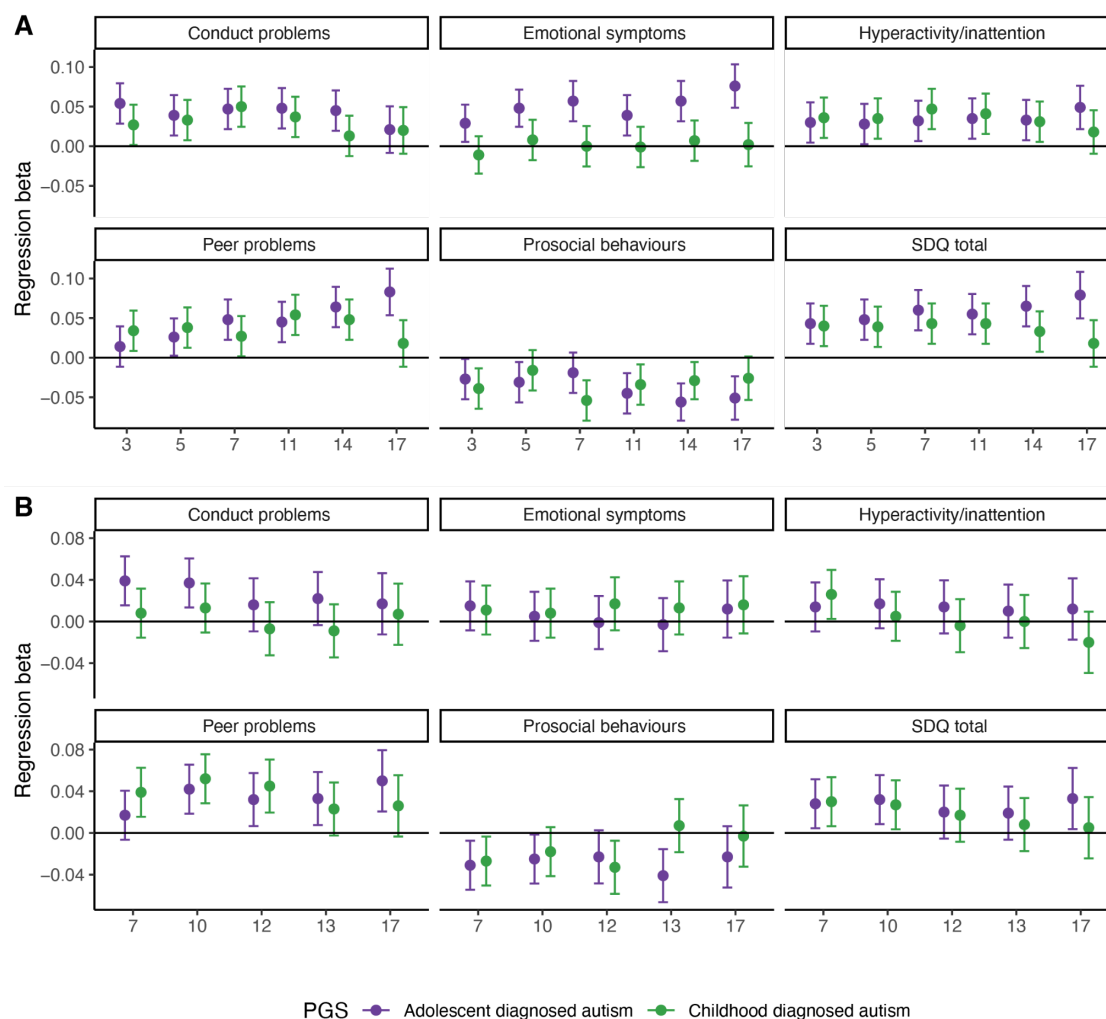


5
 6 **Extended Figure 3: SNP heritability for age at diagnosis stratified autism GWAS.** SNP
 7 heritability (points) by age at autism diagnosis for varying levels of autism prevalence. Shaded
 8 regions, 95% confidence intervals. Each vertical line indicates the best guess autism prevalence for
 9 each age at diagnosis stratified autism GWAS. SNP heritability and associated standard error (in
 10 parentheses) of autism at the best guess prevalence estimate provided in the top left corner of each
 11 facet.

12
 13 We further investigated whether genetics from age-stratified GWAS supported
 14 trajectory modelling findings. In the MCS cohort (N = 6,142 - 5,135), multivariate linear
 15 mixed effect models with age-by-PGS interaction showed that iPSYCH_{after10} PGS, but
 16 not iPSYCH_{before11}, was significantly associated with increasing emotional symptoms
 17 (BY adjusted P = 7.11×10^{-4}), peer relationship problems (BY adjusted P = 1.75×10^{-8}),
 18 SDQ total scores (BY adjusted P = 2.87×10^{-3}), and decreasing prosocial behaviours
 19 (FDR adjusted P = 0.030) with age (**Supplementary Table 18, Extended Figure 4**).
 20 Consistent results were obtained when including weights to account for sampling bias
 21 (**Supplementary Table 18**).

1 In age-by-PGS interaction analyses in ALSPAC (N = 7,172 - 4,977), $iPSYCH_{before11}$
 2 PGS was associated only with decreasing hyperactivity/inattention (BY adjusted P =
 3 4.51×10^{-3}), while $iPSYCH_{after10}$ PGS showed nominal associations with increasing peer
 4 relationship problems and SDQ total scores with age (**Supplementary Table 18**). In
 5 both cohorts, $iPSYCH_{after10}$ PGS showed larger increases in effect on SDQ total and
 6 peer relationship problems scores from childhood to adolescence compared to
 7 $iPSYCH_{before11}$ PGS. Entropy balancing did not alter these findings, suggesting that the
 8 differences in results are unlikely to be due to ascertainment differences between
 9 cohorts, but may instead reflect secular trends in mental health trajectories⁵³ or
 10 variations in the developmental periods analysed.

11



12
 13 **Extended Figure 4: Cross-sectional association between PGS for age at diagnosis stratified**
 14 **autism GWAS and socio-behavioural traits measured at different ages.** A. Association between
 15 PGS for $iPSYCH_{after10}$ and $iPSYCH_{before11}$ and scores on the SDQ total and subscales in (A) the MCS
 16 cohort at six ages (3 - 17) and (B) ALSPAC at five ages (7 - 17). For all plots, points indicate the
 17 estimate, whiskers indicate 95% confidence intervals.

18
 19 **Two correlated autism polygenic factors are associated with**
 20 **differing age at autism diagnosis**

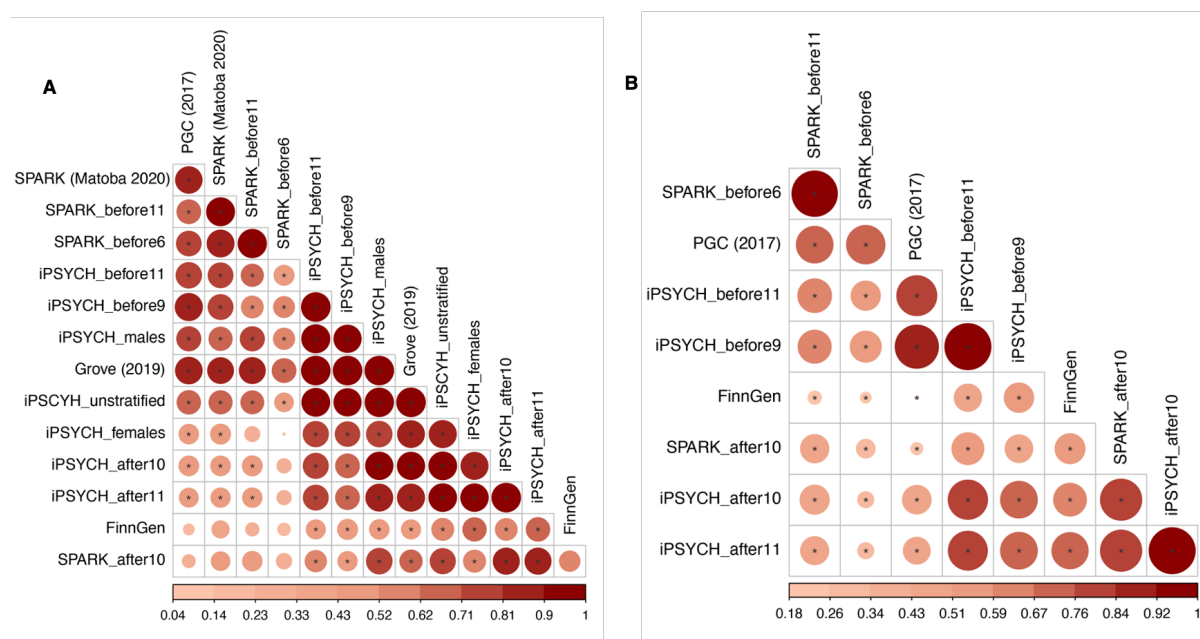
21 The above age at diagnosis stratified analyses suggested that the genetic signal
 22 underlying age at autism is likely a mixture of two or more genetic signals, with varying

1 effects on socio-behavioural trajectories. However, any age-based cutoff for diagnosis
2 is inherently arbitrary. Recognising this, we sought to understand the latent genetic
3 structures in autism using different autism GWAS and their relationship with age at
4 autism diagnosis by modelling the genetic covariances among the different autism
5 GWAS.

6
7 To enable this and provide greater resolution based on age at diagnosis, we generated
8 three additional age at diagnosis stratified GWAS of autism in SPARK using
9 (unscreened) non-autistic parents and siblings as controls ($N_{\text{control}} = 24,965$). The three
10 GWAS were: SPARK, diagnosed before age 6 (SPARK_{before6}; $N_{\text{autistic}} = 14,578$); (2)
11 SPARK, diagnosed before age 11 (SPARK_{before11}; $N_{\text{autistic}} = 18,719$); and (3) SPARK,
12 diagnosed after age 10 (SPARK_{after10}; $N_{\text{autistic}} = 3,358$).

13
14 Next, we generated genetic correlations among all the GWAS of autism we had access
15 to. We observed genetic correlations ranging from 0.04 (s.e.m = 0.14) to 0.98 (s.e.m
16 = 0.01) (**Figure 4A, Supplementary Table 19**). This was not entirely explained by
17 cohort differences or sex differences (**Supplementary Note 3**).

18
19 Hierarchical clustering of the genetic correlations identified two broad, overlapping
20 clusters (**Figure 4A**), one comprising GWAS of autism with predominantly childhood
21 diagnosed individuals, and another comprising GWAS of autism with a large fraction
22 of individuals diagnosed in adolescence or later. This pattern became clearer when
23 excluding GWAS not stratified by age at diagnosis in SPARK and iPSYCH (**Figure**
24 **4B**). This is indicative of genetic heterogeneity indexed by age at autism diagnosis.



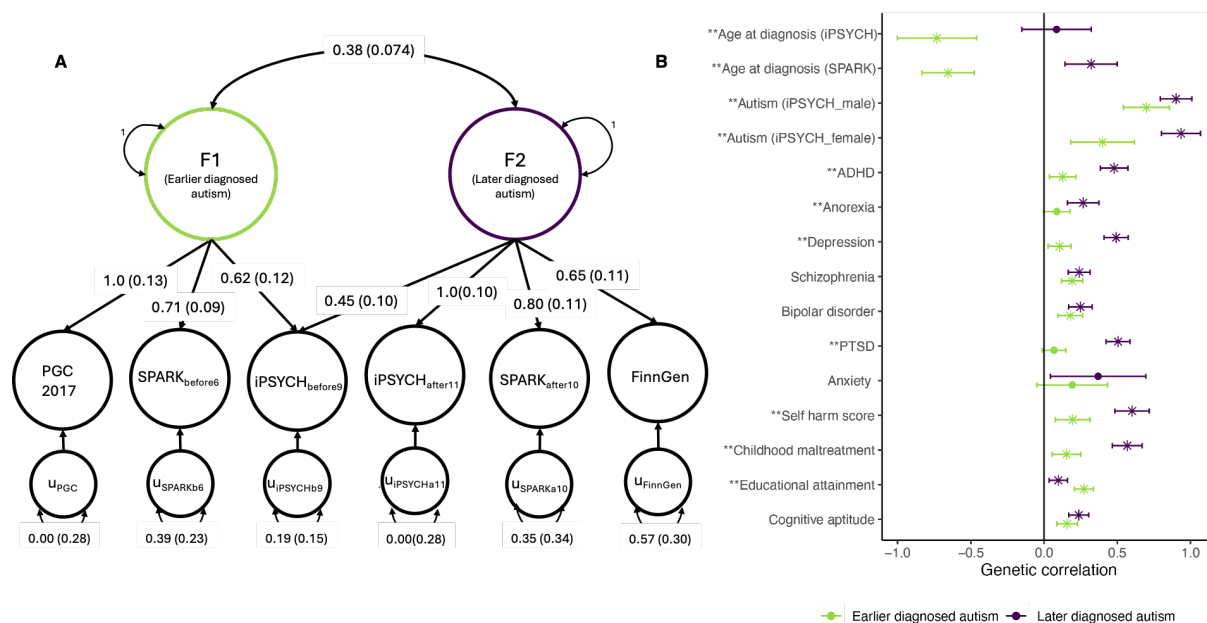
26
27 **Figure 4: Genetic correlation heatmaps between different GWAS of autism.** Genetic correlation
28 heatmaps of (A) all GWAS of autism, and (B) GWAS of autism after excluding GWAS not stratified by
29 age at diagnosis in SPARK and iPSYCH. GWAS have been ordered based on hierarchical clustering
30 of the genetic correlations. Asterisks (*) indicate significant genetic correlations after Benjamini-Yekutieli
31 adjustment.

32
33 To formally test whether the varying genetic correlation patterns among the different
34 GWAS of autism emerge from different age at diagnosis correlated latent genetic
35 traits, we modelled the genetic covariance using genomicSEM⁵⁴. GenomicSEM uses

1 structural equation models to identify latent factors. We investigated if two genetic
 2 latent traits underlie this heterogeneity, and compared it against five alternative models
 3 including a common-factor autism model (**Supplementary Table 20**). Using six
 4 minimally overlapping GWAS for autism with wide variation in age at autism diagnosis,
 5 we identified a correlated two-factor model that was the most parsimonious and fit the
 6 data best (Akaike information criterion [AIC]: 30.09, confirmatory fit index [CFI]: 1,
 7 standardised root mean residual [SRMR]: 0.039, **Figure 5A**).

8
 9 Factor 1 (Earlier diagnosed autism factor) was defined by the GWAS with
 10 predominantly early childhood diagnosed individuals (PGC-2017, SPARK_{before6}).
 11 Factor 2 (Later diagnosed autism factor) was defined primarily by GWAS with
 12 adolescent or adult diagnosed individuals (iPSYCH_{after10}, FinnGen, and SPARK_{after10}).
 13 The cross loading of iPSYCH_{before9} suggests that Factor 2 may impact behaviours in
 14 mid/late childhood as well, leading to a diagnosis before age nine. The two factors had
 15 a moderate genetic correlation ($r_g = 0.38$, s.e.m = 0.07). Factor 1 was negatively
 16 genetically correlated with both the age at autism GWAS, whilst Factor 2 positively
 17 genetically correlated with age at autism diagnosis in SPARK (**Figure 5B**), confirming
 18 that age at autism diagnosis is linked to genetic heterogeneity in autism.

19
 20 Sensitivity analyses using partly different GWAS identified a two-correlated-factor
 21 model as the best fitting model, with similar moderate genetic correlations between
 22 the two factors ($r_g = 0.39$ (0.08) - 0.52 (0.09), **Supplementary Table 20**,
 23 **Supplementary Figure 10**).



25
 26 **Figure 5: Two genetic latent factors in autism.** A. Path diagram illustrating the two-correlated-
 27 genetic-factor models for autism, using six minimally overlapping autism GWAS datasets. F1 = Factor
 28 1, F2 = Factor 2. One-headed arrows depict the regression relationship pointing from the independent
 29 variables to the dependent variables. The numbers are the regression coefficients of the factor loadings,
 30 with the standard errors provided in parentheses. Covariance between variables are represented as
 31 two-headed arrows linking the variables. The numbers on the two-headed arrows can be interpreted as
 32 genetic correlation estimates with the standard errors provided in parentheses. Residual variances are
 33 represented using a two-headed arrow connecting the residual variable (u) to itself. Standard errors are
 34 provided in parentheses. B. Genetic correlation between the two autism factors and a range of mental
 35 health, neurodevelopmental, and cognitive traits. Points indicate the estimate, whiskers indicate 95%
 36 confidence intervals, and points with an asterisk (*) indicate significant associations with Benjamini-

1 Yekutieli adjustment. Two asterisks (**) indicate phenotypes where the difference in genetic correlation
2 between earlier and later diagnosed autism is statistically significant at $P < 0.05$.

3 4 5 **Earlier and later diagnosed autism genetic factors are associated** 6 **with different mental health profiles**

7 GenomicSEM analyses revealed at least two correlated autism genetic factors. We
8 wondered if these factors are differentially genetically correlated with cognitive,
9 psychiatric and neurodevelopmental traits. Given the higher prevalence of mental
10 health diagnoses in later-diagnosed autistic individuals^{18,19}, we hypothesised this
11 might partly stem from differences in shared genetics between these autism factors
12 and other mental health and cognitive phenotypes. We tested this hypothesis using
13 genetic correlation analyses.

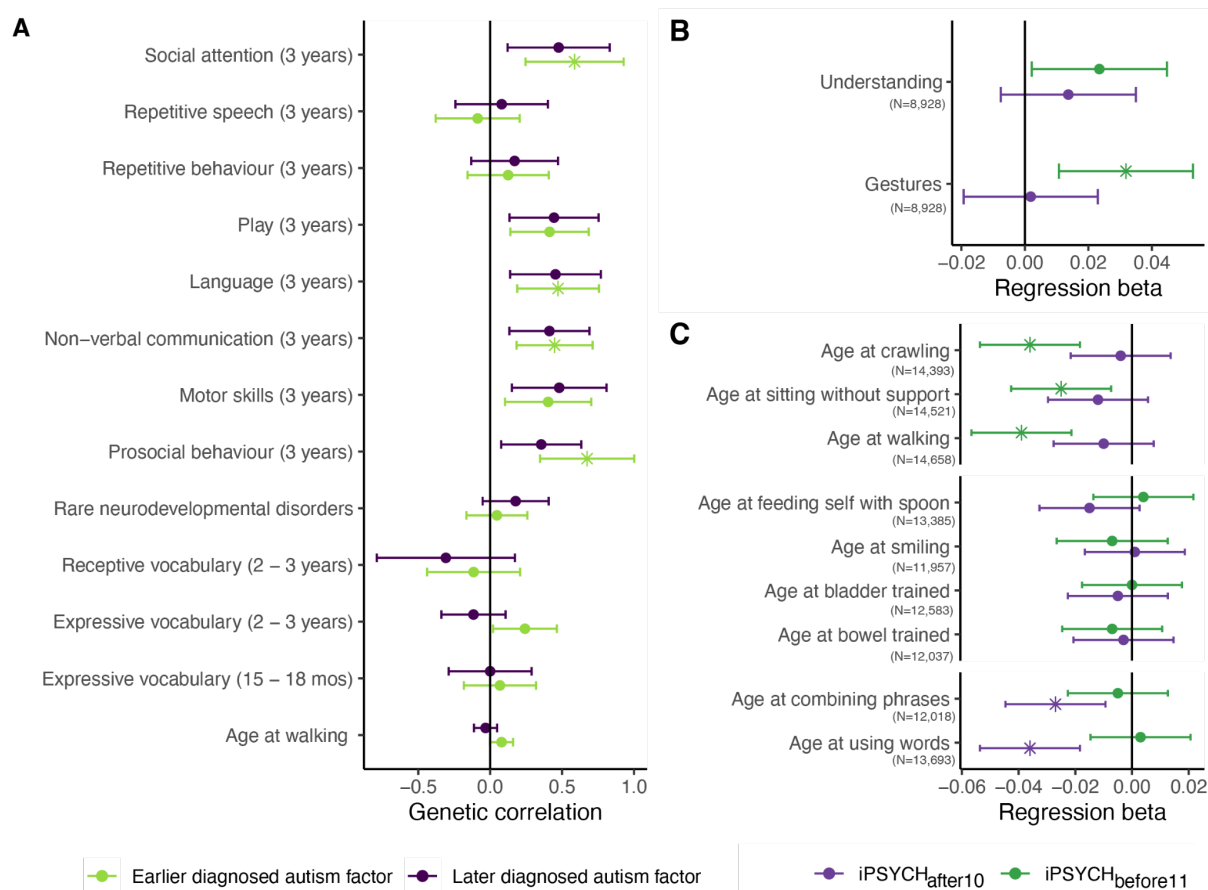
14
15 The earlier diagnosed autism factor (Factor 1) was positively correlated with
16 educational attainment and cognitive aptitude but had modest genetic correlations with
17 measures of trauma and ADHD (**Supplementary Table 21, Figure 5B**). The later
18 diagnosed autism factor (Factor 2) had lower genetic correlation with educational
19 attainment but had significant and higher positive genetically correlations with a range
20 of other mental health conditions, including internalising disorders/problems, trauma
21 and related sequelae (Depression, PTSD, childhood maltreatment, and suicide
22 attempts) and ADHD. The iPSYCH female-stratified autism GWAS (iPSYCH_{females})
23 had significantly smaller ($P = 0.028$) genetic correlation with Factor 1 than the iPSYCH
24 male-stratified autism GWAS (iPSYCH_{males}), consistent with epidemiological
25 observations that autistic females are, on average, diagnosed later than males²⁷.

26
27 Sensitivity analyses using age of diagnosis stratified GWAS from iPSYCH and SPARK
28 yielded largely consistent genetic correlation results (**Supplementary Note 6**).

29
30 We wondered if the genetic signal for later diagnosed autism can be explained by
31 diagnostic misclassification. However, decomposition of the iPSYCH autism genetic
32 signal using genomicSEM indicated that later diagnosed autism cannot be entirely
33 explained by diagnostic misclassification, although genetic effects of ADHD accounted
34 for some of the genetic variance in later diagnosed autism (**Supplementary Note 3**).
35 Accounting for ADHD's genetic effects revealed attenuated but significant moderate
36 genetic correlations between iPSYCH_{after10} and mental health conditions, suggesting
37 shared genetics with ADHD do not fully explain the elevated correlation between later
38 diagnosed autism and mental health phenotypes (**Supplementary Note 6**).

39
40 We further wondered if there is genetic overlap between the earlier diagnosed autism
41 factor and developmental milestones, as delays in developmental milestones are the
42 earliest indicator that a child may be autistic⁵⁵. The earlier but not later diagnosed
43 autism factor was positively and significantly genetically correlated with greater
44 difficulties in social behaviour at age three (**Supplementary Table 22, Figure 6A**).
45 Supporting these findings, PGS for iPSYCH_{before11} but not iPSYCH_{after10} GWAS was
46 associated with greater difficulties in social communication (gestures) at 15 months
47 (**Figure 6B, Supplementary Table 23**). Nevertheless, the genetic correlations and
48 regression coefficients did not statistically differ between the two GWAS. Furthermore,
49 PGS for neither iPSYCH_{before11} nor iPSYCH_{after10} were associated with later
50 attainment of developmental milestones among autistic individuals from SPARK.

1 However, these results may reflect collider bias as autistic carriers of rare genetic
 2 variants within the cohort have lower autism PGS and substantially delayed
 3 developmental milestones⁵⁰ (**Supplementary Table 24, Figure 6C**).
 4



5
 6
 7 **Figure 6: Association between age at diagnosis stratified autism and developmental milestones.**

8 A. Genetic correlation between earlier and later diagnosed autism genetic factors and a range of
 9 developmental phenotypes. mos = months. B. Association between PGS for iPSYCH_{before11} and
 10 iPSYCH_{after10} and social communication skills at 15 months in ALSPAC. N is the sample size. C.
 11 Association between PGS for iPSYCH_{before11} and iPSYCH_{after10} and developmental milestones among
 12 autistic individuals in SPARK. N is the sample size. For all plots, points indicate the estimate, whiskers
 13 indicate 95% confidence intervals, and points with an asterisk (*) indicates significant associations after
 14 Benjamini-Yekutieli adjustment. For all plots, positive values indicate greater difficulties/delays.
 15

16 Discussion

17 Understanding why some autistic individuals receive a diagnosis earlier than others
 18 has been an important scientific and societal question. Here we show, using multiple
 19 methods and datasets, that some of the variability in age at autism diagnosis is linked
 20 to differences in socio-behavioural trajectories and associated genetic profiles among
 21 autistic individuals. The genetic variation associated with age at autism diagnosis is
 22 also associated with the genetic heterogeneity within autism. This relationship can
 23 partly explain the often contradictory patterns of genetic correlations between autism
 24 and various cognitive, neurodevelopmental, and psychiatric phenotypes across
 25 different autism GWAS. This axis of genetic heterogeneity within autism, indexed by
 26 age at autism diagnosis, is not fully explained by several other factors that may
 27 influence age at autism diagnosis including sex, co-occurring ID and developmental

1 delays, changes to the diagnostic criteria, cohort differences, diagnostic
2 misclassification, or parental factors influencing diagnosis (**Supplementary Note 3**).

3
4 Our findings are consistent with the wider literature that demonstrates that genetic
5 influences on social-communication vary across development in the general
6 population^{37,38}. However, extending this line of enquiry, we show that genetic
7 influences on clinically diagnosed autism too vary based on age at diagnosis.
8 Modelling the genetic correlations among various GWAS of autism identified two
9 correlated autism polygenic factors that explained the data better than the alternative
10 models considered, including a single autism factor model (**Figure 5**). Notably, the
11 genetic correlation between the factors is 0.38, which is similar to the genetic
12 correlation between depression and schizophrenia⁵⁶. It is likely that other dimensions
13 contribute to heterogeneity in autism, including potentially further genetic differences
14 based on age at diagnosis. For example, a significant proportion of the variation in the
15 FinnGen autism GWAS was not explained by either of the two factors (**Figure 5A**) and
16 the correlation between some autism GWAS (e.g., iPSYCH_{females} and SPARK_{before6}) is
17 even lower than the genetic correlation observed between earlier and later diagnosed
18 autism factors.

19
20 Importantly, the low genetic correlation between the factors strongly suggests that later
21 diagnosed autism is not merely a broader autism subtype or that they emerge from
22 the same underlying latent genetic distribution. This is further supported by differences
23 in the patterns of genetic correlation between the two autism genetic factors and other
24 phenotypes, and trajectory modelling of SDQ total and subscale scores.

25
26 Of the two autism polygenic factors, we found some evidence to suggest that the
27 earlier diagnosed autism genetic factor was associated with social communication
28 difficulties in early childhood (**Figure 6**). However, neither of the two autism polygenic
29 factors are predominantly driven by a subset of children with co-occurring ID or
30 neurodevelopmental delays. The heritability of age at autism diagnosis does not
31 attenuate when accounting for ID or developmental delays, and there is no association
32 between either factor and neurodevelopmental conditions or delays in developmental
33 milestones (for example, significant delays in walking or first words). Our findings imply
34 that profound neurodevelopmental disabilities among autistic individuals may be
35 aetiologically distinct from the two autism polygenic factors.

36
37 Elevated polygenic propensity for the later diagnosed autism genetic factor may lead
38 to less clear difficulties in early childhood and thus may not be picked up by caregivers
39 as reasons to pursue diagnosis or support. These are consistent with findings that
40 children who do not initially meet the criteria for an autism diagnosis may later meet
41 them^{7,10}, and with parental reports of on average “milder” autism features among later
42 diagnosed autistic individuals⁵⁷.

43
44 In contrast, both the later diagnosed genetic autism factor (**Figure 5**) and the late
45 emergent latent class of SDQ total scores (**Figure 1**) are associated with greater
46 mental health problems, particularly internalising difficulties, self-harm, and correlates
47 of trauma. The later diagnosed autism genetic factor is also associated with a larger
48 increase in socio-behavioural difficulties in adolescence. Yet again, this is consistent
49 with the epidemiological findings of greater mental health difficulties among later
50 diagnosed autistic individuals^{18,19}. However, our findings demonstrate that part of the

1 epidemiological findings may be explained by genetic heterogeneity in autism indexed
2 by age at autism diagnosis. How exactly these genetic differences lead to greater
3 mental health problems remains to be resolved.

4
5 The variation in genetic correlation between ADHD and autism stratified by age of
6 diagnosis is particularly noteworthy. Older GWAS of autism (including the PGC-2017)
7 were not significantly genetically correlated with ADHD^{51,58,59} whereas more recent
8 GWAS for autism have moderate genetic correlations with ADHD⁶⁰. Genetic
9 correlation analyses (**Figure 5**) indicate that the genetic correlation with ADHD
10 increases with a later diagnosis of autism. We confirmed this using within-family
11 analyses: autistic individuals diagnosed before age five do not over-inherit PGS for
12 ADHD (**Supplementary Table 13c**). However, even within ADHD, there is genetic
13 heterogeneity based on age at diagnosis⁶¹. ADHD diagnosed in childhood shows a
14 larger genetic correlation with autism and lower genetic correlation with internalising
15 disorders⁶² compared to late-diagnosed ADHD. These results suggest a complex
16 genetic interplay between autism and ADHD that is dependent on age at diagnosis.

17
18 Several sensitivity analyses indicate that our findings are not primarily capturing sex
19 differences (**Supplementary Note 3**). However, given that autistic females are, on
20 average, diagnosed later than males^{4,5}, research that investigates sex and gender
21 differences in both autism and co-occurring conditions⁶³ needs to account for genetic
22 confounding associated with age at autism diagnosis. Findings that may be thought to
23 reflect sex differences may additionally reflect differences in age at diagnosis. For
24 example, the higher prevalence of mental health problems in autistic females
25 compared to males⁶⁴ attenuates when restricting to autistic individuals diagnosed
26 before age five¹⁸.

27
28 Although we explored the impact of several additional clinical and demographic factors
29 on age at autism diagnosis, these account for less than 50% of the variance in age at
30 diagnosis (**Supplementary Table 8** and **Supplementary Note 5**). Notably, there is
31 substantial variation across the datasets explored, highlighting that age at diagnosis
32 of autism is immensely complex, varying across geography and time. Local cultural
33 factors, access to healthcare, gender bias, stigma, and camouflaging, all of which are
34 difficult to measure, likely have an impact on who receives a diagnosis and when.

35
36 Of interest is camouflaging, which has been hypothesised as one reason for later
37 diagnosis⁶⁵ particularly among autistic females. Our results cannot be fully explained
38 by camouflaging. Although children as young as seven years of age may
39 camouflage⁶⁶, it is unlikely that infants can camouflage behaviours or developmental
40 milestones in infancy or early toddlerhood. Furthermore, even among autistic
41 individuals, there is variation in camouflaging^{67,68}. Although we are unable to explicitly
42 test the impact of camouflaging, our results are consistent with the correlates of
43 camouflaging. For instance, it is known that autistic females are more likely to
44 camouflage⁶⁹, and higher camouflaging is associated with greater mental health
45 difficulties⁷⁰, and later autism diagnosis^{65,71}.

46
47 In conclusion, using genetic data and longitudinal analyses of birth cohorts, we identify
48 an axis of heterogeneity in autism which is indexed by age at autism diagnosis. This
49 axis of heterogeneity partly explains the varying genetic correlations among the
50 different GWAS of autism and between autism and various mental health conditions.

1 These findings provide further support to the hypothesis that the umbrella term
2 “autism” describes multiple phenomena with differing aetiologies, developmental
3 trajectories, and correlations with mental health conditions. These findings have
4 implications for how we conceptualise neurodevelopment more broadly, and for
5 understanding diagnosis, sex and gender differences, and co-occurring health profiles
6 in autism.

1 **Methods**

2 **A note on terminology**

3 We use the term autistic to refer to people who have an autism diagnosis⁷². We use
4 non-autistic to refer to people who do not have an autism diagnosis. We use sex to
5 refer to sex assigned at birth, and use the terms males and females to refer to sex.
6

7 **Analyses of birth cohorts**

8 **Study design and participants**

9 We used four population-based birth cohorts that vary in ages covered and when data
10 were collected (**Extended Figure 1**). Briefly, the four cohorts included are the UK-
11 based Millennium Cohort Study (MCS)⁷³, the Ireland-based Growing Up in Ireland
12 (GUI) Child cohort (aka Cohort 98')⁷⁴, and the Australia-based Longitudinal Study of
13 Australian Children - Birth (LSAC-B) and Kindergarten (LSAC-K) cohorts⁷⁵. All children
14 included in the cohorts were born in the 21st century. Further details about the cohorts
15 are provided in the **Supplementary Note 1**.
16

17 As indicated in **Supplementary Table 1**, these cohorts were adopted for their
18 longitudinal nature, the nationally representative cohort members, and the availability
19 of data on behavioural profiles and neurodevelopmental diagnosis allowing for cross-
20 country comparisons and generalisation⁷⁶.
21

22 **Measures**

23 *Autism and ADHD diagnosis and age at diagnosis*

24 In all cohorts, across multiple sweeps, the main caregiver was asked if the participant
25 had a diagnosis of autism (**Extended Figure 1**). For age at diagnosis, we used the
26 age at the sweep when participants first reported being diagnosed as autistic in every
27 cohort, to maximise sample sizes and ensure consistency across cohorts for effective
28 comparisons. For instance, if a participant first reported an autism diagnosis at the age
29 11 sweep, we considered age at diagnosis to be 11 years. Although the specific age
30 at diagnosis was provided for LSAC-B and LSAC-K, we opted not to use this, as we
31 identified errors in some reports where months and years of diagnosis were swapped
32 or not reported.
33

34 In MCS, we had reports of both autism and ADHD diagnoses, allowing us to conduct
35 several sensitivity analyses. For our primary analyses, we included a narrowly defined
36 sample of children with consistently reported autism diagnoses by primary and proxy
37 caregivers (when both were available) and no other reported neurodevelopmental
38 diagnosis (particularly ADHD). To assess the generalizability of our results and
39 increase the sample size, we then expanded the sample to include all children with
40 any reported diagnosis of autism. This expanded sample, which we refer to as "MCS-
41 expanded", included cases regardless of whether the diagnoses were consistent
42 across sweeps or caregivers, and included those with co-occurring ADHD.
43 Additionally, we imputed the independent variables and covariates for autistic
44 individuals with missing information, as detailed below. We refer to this sample as
45 "MCS-imputed". Finally, to assess the specificity of the trajectories for autism, we
46 conducted analyses among children who had a consistent ADHD diagnosis but no
47 diagnosis of autism. We refer to this sample as "MCS-ADHD".
48

1 *Strengths and Difficulties Questionnaire (SDQ)*

2 We used the SDQ to capture social, emotional, and behavioural profiles of
3 participants, with repeated measures from 3 to 18 years across cohorts (**Extended**
4 **Figure 1**). SDQ comprises 25 statements that respondents are asked to rate on a 3-
5 point Likert scale (“not true”, “somewhat true”, and “certainly true”) based on the child’s
6 symptoms or behaviours over the past six months. There are five subscales, each
7 containing five items, which assess emotional symptoms, conduct problems,
8 hyperactivity-inattention, peer relationship problems, and prosocial behaviours
9 respectively²⁰. The first four subscales assess difficulties, and the total score ranges
10 from 0 to 40, with higher scores indicating more significant difficulties. The fifth
11 subscale represents strengths and has a total score ranging from 0 to 10, with higher
12 scores indicating more prosocial behaviours. The SDQ demonstrates good test-retest
13 reliability and criterion validity across countries^{21–23}. Each subscale of SDQ has been
14 found to exhibit correlations with diagnosis of autism and ADHD⁷⁷. Its five-factor
15 structure (each subscale as a factor) has shown consistency and invariance across
16 age, sex, and ethnic background^{21,25}. The SDQ captures several core features of
17 mental health and neurodevelopmental conditions, including autism and ADHD⁷⁸. Only
18 children with complete data of SDQ across all sweeps were included in the analyses,
19 except for imputation analyses.

20 *Socio-demographic measures*

21 Socio-demographic measures were included as covariates to account for their impact
22 on age at diagnosis in each cohort (**Supplementary Table 25**). Specific measures
23 and available information vary across cohorts, but we generally included sex, ethnic
24 background, maternal age at delivery, child’s cognitive aptitude, household socio-
25 economic status (SES), and deprivation level of the living area, to account for some
26 factors that may impact the age when someone receives an autism diagnosis^{32,79}. Only
27 subsets of children in the complete-SDQ samples, with complete data for these socio-
28 demographic factors, were included in the respective analyses, resulting in a further
29 reduction in sample sizes.

30
31
32 In MCS, although various census classifications for ethnic groups were available, we
33 opted to use a binary indicator to identify non-white ethnic minorities. This approach
34 was chosen to maintain consistency with other cohorts. Ethnicity data were not
35 collected in either LSAC cohort. Instead, visible ethnic minority status was determined
36 primarily by parental country of birth and the language(s) spoken at home.⁸⁰ Maternal
37 age at delivery was collected only in MCS. In other cohorts, we used maternal age (in
38 years) at first sweep of data collection to reflect the variation in maternal age at
39 delivery.

40
41 For cognitive aptitudes and other socio-demographic factors, including SES and
42 deprivation, we adopted summary scores using principal component analysis (PCA)
43 to capture information measured by diverse scales. However, as including more social
44 factors lead to smaller sample sizes, we prioritised factors based on the availability of
45 data among already limited autistic samples. Information on socio-demographic
46 factors in each cohort, including variables and scales included in PCA and resulting
47 sample sizes, can be found in **Supplementary Table 25**. Note, in the birth cohorts, no
48 autistic child was identified as having intellectual disability (ID), defined as scoring two
49 or more standard deviations below the mean value of the first principal component

1 score ('g' factor) derived from multiple cognitive aptitude measures in corresponding
2 cohort.

3

4 **Statistical analyses**

5 Following the participant selection process, we used two methods to model the
6 longitudinal trajectories of SDQ total and subscale scores. In the first analyses, we a
7 *priori* defined two groups of autistic individuals - childhood diagnosed (diagnosed
8 before ages 9 - 11, depending on the cohort), and adolescent diagnosed (diagnosed
9 after the ages of 9 - 11, depending on the cohort). This age period was chosen as
10 there is epidemiological evidence showing increased autism incidence among females
11 during this window²⁷ and because trajectory analyses have identified increasing
12 autism-related traits in a subset of individuals after this period⁸¹. We were also limited
13 in choosing alternate cutoffs due to the absence of information on both SDQ and
14 autism diagnosis at earlier and later time points in some of the cohorts, and the
15 relatively low sample sizes of the resulting groups.

16

17 Anyone who had no report of an autism diagnosis were included in the general
18 population group. For MCS in particular, children with neither autism nor ADHD
19 diagnosis were included in the general population group. We used linear Latent
20 Growth Curve Models (LGCM) to identify the latent trajectories of SDQ total and
21 subscale scores in the three groups (childhood diagnosed, adolescent diagnosed, and
22 the general population) for all cohorts. Each linear model included a latent intercept to
23 represent the initial level of the outcome variable, and a linear latent slope to represent
24 the mean rate of change over time. As sensitivity analyses, quadratic models were
25 also fitted in MCS, MCS-expanded, and LSAC-B, in which quadratic time scores were
26 assigned across sweeps respectively to capture this nonlinear change over time.
27 However, quadratic models for most subscales among the three cohorts either did not
28 converge or demonstrated Heywood cases, i.e., negative variance estimates, in slope
29 terms This likely stems from insufficient statistical power due to the small sample sizes
30 or model misspecification, which hinders meaningful theoretical interpretation.
31 Therefore, we decided to use linear models.

32

33 Given the well-known sex differences in diagnosis age²⁷, we also applied the same
34 models stratified by sex, i.e., estimating latent intercept and slope for each sex, within
35 the autistic samples. All LGCM were fitted under the structural equation modelling
36 framework using the *lavaan* package in R⁸².

37

38 In parallel, we conducted Growth Mixture Models (GMM) to identify if there were latent
39 groups of autistic individuals based on their trajectories of SDQ total and subscale
40 scores. GMM assumes that the sample being studied consists of multiple mixed
41 effects models, each capturing a subgroup trajectory with shared intercept and slope⁸³.
42 We fitted models with one to four groups for each subscale and SDQ total scores in
43 each cohort, using the *lcmm* package in R⁸⁴. The optimal number of latent trajectories
44 were then determined by comparing fit indices, including Bayesian Information
45 Criterion (BIC) values, classification quality measure (entropy), and substantive
46 interpretation. Models with lower BIC values and higher entropy are favoured⁸⁵. Also,
47 models identifying subgroups with less than 5% of the sample size were not
48 considered for poor statistical reliability and limited practical significance⁸⁶.

49

1 Multiple regressions were subsequently conducted to investigate the association
2 between individuals' age at diagnosis (the outcome variable) and their SDQ total and
3 subscale latent class memberships identified in optimal GMM, as well as other socio-
4 demographic covariates. We did not detect any multicollinearity among the variables
5 using variance inflation factors. Estimates of coefficients of predictors and
6 corresponding p-values were interpreted to determine which factors contribute to
7 differing age at autism diagnosis.

8
9 Additionally, considering the limited sample sizes and the number of explanatory
10 variables included, the relative importance of each predictor was assessed using
11 dominance analysis⁸⁷. We employed the *misty*⁸⁸ package in R for this analysis, using
12 a correlation matrix extracted from the fitted model via the *lavInspect* function from the
13 *lavaan* package⁸². This approach leverages the correlation matrix to consider not only
14 individual predictors but also the correlations among them, providing a more
15 comprehensive assessment of their relative importance⁸⁹.

16
17 To examine potential causal pathways, mediation analyses were conducted, allowing
18 socio-demographic factors to indirectly influence the age at diagnosis through their
19 effects on latent class memberships identified in the optimal GMM. Using structural
20 equation modelling in the *lavaan* package⁸², both direct and indirect effects were
21 assessed, with their significance calculated using bootstrapping analysis. Further
22 details are provided in **Supplementary Note 5**.

23
24 To investigate the specificity of our findings to autism, we conducted LGCM, GMM,
25 and sequential mediation analyses in individuals with ADHD but without a co-occurring
26 autism diagnosis in the MCS cohort (N = 89, **Supplementary Table 4**). ADHD
27 diagnoses were available in the same sweeps as autism diagnoses, reported at age
28 5,7,11, and 14. Carers were asked the following question: 'Has a doctor or other health
29 professional ever told you that <child's name> had Attention Deficit Hyperactivity
30 Disorder (ADHD)?'.

31
32 In autistic individuals, using the GMM based latent classes of the SDQ total scores,
33 we used multiple regression to investigate the association with mental health
34 phenotypes in MCS, LSAC-K, and LSAC-B. We included sex as a covariate.

35 36 *Imputation*

37 To assess the impact of missingness, we applied *softImpute*, to impute missing data
38 for all children with an autism diagnosis reported by any carer in any sweep in the MCS
39 cohort (N = 623, **Supplementary Table 3**). Given the longitudinal nature of data
40 collections for SDQ subscale scores and some cognitive aptitude measures,
41 *SoftImpute* was chosen for its computational efficiency in handling large-scale
42 matrices through low-rank approximation, effectively preserving underlying structure
43 of input data. To enhance imputation quality and reduce bias, we included related
44 auxiliary variables in the imputation process, along with SDQ subscale scores in all
45 available sweeps (**Supplementary Table 3**). Further information is provided in
46 **Supplementary Note 2**.

47 48 **SPARK cohort: Genotyping, quality control and imputation**

49 We used data from the SPARK cohort³⁹ iWES2 v1 dataset (released in Feb 2022)
50 which included data from 70,487 autistic individuals and their families. All participants

1 were genotyped using the Illumina Global Screening Array (GSA_24v2-0_A2). To
2 avoid false positives due to fine-scale population stratification, we restricted the
3 analyses to individuals of genetically-inferred European ancestries (N = 51,869 autistic
4 and non-autistic participants), which was provided by the SPARK consortium. From
5 this, we restricted to individuals with genotyping rate > 98%, individuals without sex
6 mismatches and excess heterozygosity (3 standard deviations from the mean
7 heterozygosity), and where trio data was available, trios with fewer than 5% Mendelian
8 errors, resulting in 47,170 autistic and non-autistic individuals. We included genetic
9 variants with minor allele frequency > 1%, genotyping rate > 95%, and that were in
10 Hardy Weinberg Equilibrium (HWE p-value > 1×10^{-6}), resulting in 518,189 SNPs.

11
12 We used this quality controlled genotype data for imputation, calculating genetic
13 principal components, and inferring relatedness among individuals. We inferred
14 genetic relatedness using KING⁹⁰. For genetic principal component analysis, we
15 pruned SNPs for linkage disequilibrium (LD) (maximum $r^2 = 0.1$) and removed the
16 human leukocyte antigen (HLA) region. Using PC-AiR⁹¹, we first calculated principal
17 components (PCs) in genetically unrelated individuals and then projected the PCs onto
18 related individuals. We imputed genotypes using the TOPMED imputation panel⁹² on
19 the Michigan imputation server (v1.7.3)⁹³ using Minimac4⁹³ and after phasing using
20 Eagle v2.5⁹⁴. Post imputation, variants were converted from GRCh38/hg38 to
21 GRCh37/hg19 using liftOver. We restricted downstream analyses only to variants with
22 minor allele frequency > 0.1% and with an imputation $R^2 > 0.6$.

23 24 25 **SPARK cohort: Association analyses**

26 *PGS association analyses*

27 Polygenic scores (PGS) were calculated using PRSCs⁹⁵ which uses a Bayesian
28 shrinkage prior. PGS were calculated for autism (iPSYCH only dataset, N = 19,870
29 autistic individuals and 39,078 non-autistic individuals)^{50,96}, ADHD⁴⁰, bipolar
30 disorder⁹⁷, major depressive disorder⁴⁴, schizophrenia⁴¹, educational attainment⁹⁸,
31 and cognitive aptitude⁴⁶, autism diagnosed before age 11 (iPSYCH_{before11}), and autism
32 diagnosed after age 10 (iPSYCH_{after10}). The latter two GWAS were generated using
33 the iPSYCH2015⁹⁶ cohort, details of which are provided below. For simplicity we refer
34 to this cohort as iPSYCH throughout.

35
36 We ran association analyses between PGS and age at autism diagnosis (converted
37 to years in all analyses) in the quality controlled dataset. We excluded individuals older
38 than 22 to focus on those who had an autism diagnosis using either the DSM-IV⁹⁹ or
39 DSM-5, retaining a maximum of 18,809 autistic individuals for PGS analyses. This
40 criteria also allowed us to focus on individuals who received their diagnosis in
41 childhood or adolescence, as older adults may have missed an earlier diagnosis of
42 autism due to secular changes in social attitudes towards autism. For psychiatric
43 conditions, we ran separate linear regressions with age at autism diagnosis and the
44 aforementioned PGS. The baseline model included ID (caregiver reported), sex, and
45 the first 10 genetic principal components as covariates. For schizophrenia, ADHD,
46 depression, cognitive aptitude, educational attainment, iPSYCH_{before11} and
47 iPSYCH_{after10}, we ran sensitivity analyses by sequentially including age at walking and
48 age at first words (developmental milestones), parental occupation, highest parental
49 education, and household income (together, socio-economic status or SES), and
50 national area deprivation percentile (deprivation) as covariates. We also included trio

1 status in the baseline model to account for potential participation bias. Additionally, for
2 $iPSYCH_{before11}$ and $iPSYCH_{after10}$ we included any diagnosis of an attentional or
3 behavioural disorder as a covariate in the baseline model to account for diagnostic
4 overshadowing. For the PGS with $iPSYCH_{before11}$ and $iPSYCH_{after10}$, we ran sensitivity
5 analyses after stratifying by sex.

6
7 We tested if the effects varied by sex by including a PGS by sex interaction term in the
8 baseline model.

9
10 We tested for direct and indirect genetic effects of ADHD and educational attainment
11 (EA) PGS in two ways. First, we generated pseudocontrols for complete trios in Plink
12 1.9^{100,101}, and calculated PGS separately for autistic individuals and pseudocontrols.
13 We regressed the effects of the ADHD and EA PGS for autistic individuals and
14 pseudocontrols on age at autism diagnosis. We included sex, ID (caregiver reported),
15 and the first 10 genetic principal components as covariates. The direct effects were
16 calculated by subtracting the effects of the untransmitted PGS (indirect effect) from
17 the transmitted PGS (total effects). Standard errors and p-values were calculated by
18 bootstrapping 10,000 times as done previously¹⁰².

19
20 Second, in complete trios, we used polygenic transmission disequilibrium tests
21 (pTDT)¹⁰³ to calculate the deviation of PGS from the parental mean PGS for ADHD
22 and EA. We checked for over-transmission of PGS stratified by sextiles based on age
23 at autism diagnosis.

24
25 We used pTDT to investigate if there is an over-transmission of PGS for $iPSYCH_{before11}$
26 and $iPSYCH_{after10}$ among autistic individuals in the SPARK cohort.

27
28 In the SPARK cohort, we obtained data for age at achieving nine developmental
29 milestones (in months) for autistic individuals. For all milestones, we excluded
30 individuals who were greater than five median absolute deviations from the median.
31 We ran multiple linear regression with PGS for $iPSYCH_{before11}$ and $iPSYCH_{after10}$ in
32 which we included sex, age at recruitment into the study, and the first 10 genetic
33 principal components as covariates.

34 35 *Rare high-impact de novo variants and inherited variants*

36 We identified rare (minor allele frequency [MAF] < 0.1%) *de novo* and inherited
37 variants in complete trios from SPARK as previously described¹⁰⁴. We identified high
38 impact protein truncating variants by restricting variants in loss-of-function
39 observed/expected upper bound fraction (LOEUF)¹⁰⁵ highly constrained decile
40 (LOEUF < 0.37) that were annotated as either frameshift, stop gained, or start lost;
41 and had a loss-of-function transcript effect estimator (LOFTEE) “high confidence
42 annotation”. To identify high-impact *de novo* missense variants, we restricted to
43 variants in LOEUF highly constrained genes (LOEUF < 0.37), and had an MPC
44 (missense badness, PolyPhen-2, and constraint) score¹⁰⁶ > 2. All variants were rare,
45 with an allele frequency < 0.1% in SPARK and gnomAD.

46
47 We ran regression analyses separately for high-impact *de novo* and inherited protein
48 truncating variants and missense variants, and additionally by combining both protein
49 truncating and missense variants. We included sex and age at recruitment into the

1 study as covariates for analyses with *de novo* variants, and additionally the first 10
2 genetic principal components for analyses with inherited variants.

4 **GWAS for age at autism diagnosis and age at diagnosis stratified autism**

5 We generated a GWAS of age at autism diagnosis (in years) on the quality controlled
6 dataset from SPARK, restricting it to individuals who were under 22 years of age ($N =$
7 18,809), and SNPs with a MAF > 1%. GWAS was generated using FastGWA¹⁰⁷ with
8 age at recruitment into the study, sex, ID, and the first 10 genetic principal components
9 as covariates. In iPSYCH, we generated an additional GWAS of age at autism
10 diagnosis (in years) in a quality controlled dataset of unrelated individuals with sex and
11 ID included as covariates using FastGWA¹⁰⁷, restricting to SNPs with an MAF > 1%.
12 To keep it consistent with SPARK, we excluded individuals who were diagnosed after
13 age 22, resulting in 18,965 individuals. Briefly, pre-imputation quality control of the
14 iPSYCH data was performed using the Ricopili pipeline¹⁰⁸, prephased using Eagle
15 v.2.3.5, and imputed using Minimac3¹⁰⁹, using the downloadable version of the
16 Haplotype Reference Consortium (HRC)¹¹⁰ (accession no. EGAD00001002729).
17 Further details of quality control and imputation are provided in Als et al., 2023⁴⁴.

18
19 We additionally generated three age at autism diagnosis stratified GWAS in SPARK
20 using (unscreened) non-autistic parents and siblings as controls ($N_{\text{control}} = 24,965$).
21 The three GWAS were: (1) SPARK, diagnosed before age 6 (SPARK_{before6}; $N_{\text{autistic}} =$
22 14,578); (2) SPARK, diagnosed before age 11 (SPARK_{before11}, $N_{\text{autistic}} = 18,719$); and
23 (3) SPARK, diagnosed after age 10 (SPARK_{after10}, $N_{\text{autistic}} = 3,358$). For these analyses,
24 we did not restrict it to individuals under 22 to increase sample size. GWAS were
25 generated using quality controlled SNPs with a MAF > 1% using FastGWA-GLMM¹¹¹.
26 We included age at recruitment into the study (to account for parental controls
27 potentially lacking autism diagnoses due to historical diagnostic changes), sex and the
28 first 10 genetic principal components as covariates. Fast-GWA GLMM can account for
29 relatedness and fine-scale population stratification even in family-based samples like
30 SPARK.

31
32 Although inclusion of unscreened related individuals as controls can decrease
33 heritability and statistical power to identify loci¹¹², we used the GWAS to primarily
34 conduct genetic correlation and related analyses. To ensure the robustness of these
35 models we: (1) confirmed that the attenuation ratio for all GWAS was not significantly
36 greater than 1; (2) generated an additional GWAS of SPARK without stratifying by age
37 at autism diagnosis using the same methods and confirmed a high genetic correlation
38 ($r_g = 0.92$, s.e.m = 0.17) with a previous SPARK GWAS⁵² which used a case-
39 pseudocontrol approach; and (3) in the genomicSEM analyses, ran sensitivity
40 analyses using a trio-based SPARK GWAS in lieu of the age at diagnosis stratified
41 GWAS from SPARK and confirmed our findings.

42
43 We generated four age at diagnosis stratified GWAS of autism in iPSYCH cohort⁹⁶.
44 The primary GWAS used in the analyses were GWAS of autism diagnosed before age
45 11 (iPSYCH_{before11}: 9,500 autistic and 36,667 non-autistic individuals) individuals and
46 autism diagnosed after age 10 (iPSYCH_{after10}: 9,231 autistic and 36,667 non-autistic
47 individuals). We chose this age cutoff to divide the iPSYCH cohort into two subgroups
48 with similar sample sizes and because age coincided with the window in which we
49 observe an increase in SDQ scores in the birth cohorts, and which is associated with
50 an increase in diagnosis of females in epidemiological samples²⁷.

1
2 We conducted two additional GWAS with smaller sample sizes: GWAS of autism
3 diagnosed before age nine (iPSYCH_{before9}: 5,451 autistic and 36,667 non-autistic
4 individuals) and after age 11 (iPSYCH_{after11}: 8,260 autistic and 36,667 non-autistic
5 individuals). These were used in sensitivity analyses. For the last two GWAS, we also
6 conducted GWAS after excluding individuals born after 1994, to ensure that all autistic
7 individuals received a diagnosis using either DSM-IV or DSM-5 criteria. However, we
8 observed high genetic correlation between the GWAS when using the full sample and
9 when restricting the sample to those born after 1994, suggesting that changes in the
10 diagnostic criteria do not substantially impact the genetic analyses. To increase
11 sample size and statistical power, we conducted all downstream analyses without
12 excluding autistic individuals born before 1994.

13
14 All individuals included in these GWAS from iPSYCH were born between May 1980
15 and December 2008 to mothers who were living in Denmark. GWAS was conducted
16 on individuals of European ancestry, with the first 10 genetic principal components
17 included as covariates using logistic regression as provided in PLINK.

18 **Heritability, genetic correlation, and genomicSEM**

19 Heritability analyses for age at autism diagnosis were conducted using a single-
20 component genome-wide complex trait analysis with genomic-relatedness-based
21 restricted maximum likelihood approach (GCTA-GREML)^{113,114} in unrelated autistic
22 individuals using the quality controlled genetic data in SPARK. We estimated SNP-
23 based heritability first after including sex, age, and the first ten genetic principal
24 components as covariates. We ran sensitivity analyses after sequentially including ID,
25 developmental milestones, SES and area deprivation as covariates.

26
27
28 SNP-based heritability for the age at diagnosis stratified autism GWAS from iPSYCH
29 was calculated using linkage disequilibrium score regression coefficient (LDSC)^{58,115}
30 using linkage disequilibrium scores from the north-west European population. We
31 converted observed scale heritability estimates to liability scale estimates using a
32 range of autism lifetime prevalence estimates, including a “best guess” autism lifetime
33 prevalence estimate for each of the age-stratified autism GWAS.

34
35 We conducted genetic correlation analyses using LDSC, using linkage disequilibrium
36 scores from the north-west European populations.

37
38 For genomicSEM⁵⁴ analyses, we first conducted genetic correlation analyses among
39 fourteen different autism GWAS using LDSC. This included a multi-ancestry case-
40 pseudocontrol GWAS in SPARK⁵² (6,222 case-pseudocontrol pairs); GWAS from
41 FinnGen (Data Release - r10)¹¹⁶ (646 cases and 301,879 controls), the PGC-2017
42 autism GWAS⁵¹ (7,387 cases and 8,567 controls), seven GWAS from iPSYCH, and
43 age at diagnosis stratified GWAS from SPARK. The iPSYCH GWAS included an
44 unstratified (19,870 autistic individuals [15,025 males and 4,845 females] and 39,078
45 controls) and sex-stratified GWAS⁵⁰, and four age at diagnosis stratified GWAS as
46 mentioned earlier.

47
48 Subsequently, we restricted to six GWAS with minimal sample overlap, without high
49 genetic correlation ($r_g > 0.95$), and with wide variation in age at diagnosis to conduct
50 genomicSEM analyses using autosomes. Using the patterns of genetic correlations

1 observed we tested an age at diagnosis related correlated two-factor model. We
2 additionally tested: (1) a single-factor model; (2) a correlated two-factor “geography”
3 model where three US-based autism GWAS loaded onto one factor, and three Europe-
4 based autism GWAS loaded onto a second factor; (3) a bifactor model based on age
5 at diagnosis; (4) a bifactor model based on the geography of the cohorts; and (5) a
6 hierarchical factor model based on age at diagnosis. The two-factor model was chosen
7 as it had lower RMSEA and higher CFI and was more parsimonious than the bifactor
8 model. We ran sensitivity analyses using different GWAS of autism as input and
9 confirmed that the two-correlated-factor model was the best fitting model of the models
10 tested.

11
12

13 **Analyses in ALSPAC and MCS**

14 *Genetic quality control*

15 We obtained quality controlled and imputed genotype data from ALSPAC^{117–119}.
16 Further details about the cohort are provided in the **Supplementary Note 1**. Briefly,
17 ALSPAC children were genotyped using the Illumina HumanHap550 quad chip
18 genotyping platforms by 23andme. Individuals were excluded due to sex mismatches,
19 excess heterozygosity, missingness > 3%, and insufficient sample replication
20 (Identical-By-Descent [IBD] < 0.8). After multidimensional scaling, and comparison
21 with Hapmap II (release 22), only individuals of genetically inferred European
22 ancestries were retained. SNPs with low frequency (MAF < 1%), poor genotyping (call
23 rate < 95%) and deviations from Hardy-Weinberg equilibrium ($P < 5 \times 10^{-7}$) were
24 removed. 9,115 subjects and 500,527 SNPs passed quality control. Genotypes were
25 phased using ShapeIT, and imputation was done using the Haplotype Reference
26 Consortium panel using the Michigan imputation server. After imputation, we further
27 removed low frequency SNPs (MAF < 1%). Further details of the quality control and
28 imputation of ALSPAC are provided here:
29 https://proposals.epi.bristol.ac.uk/alspac_omics_data_catalogue.html#org89bb79b.

30 Genome-wide genotype data was generated by Sample Logistics and Genotyping
31 Facilities at Wellcome Sanger Institute and LabCorp (Laboratory Corporation of
32 America) using support from 23andMe.

33

34 We also obtained quality controlled and imputed data from MCS. Briefly, MCS samples
35 were genotyped using the Illumina Global Screening Array¹²⁰. Individuals were
36 excluded due to sex mismatches, excess heterozygosity, and missingness > 2%. We
37 identified European samples using the GenoPred pipeline¹²¹
38 (<https://github.com/opain/GenoPred>). SNPs with low frequency (MAF < 1%), poor
39 genotyping (call rate < 97%) and deviations from Hardy-Weinberg equilibrium ($p < 1 \times 10^{-6}$)
40 were removed. Imputation was conducted using Minimac4⁹³ using the
41 TOPMED reference panel⁹² in the Michigan imputation server⁹³. Post imputation,
42 SNPs with an imputation R^2 INFO score < 0.8, with > 3% missing, and with a MAF <
43 1% were excluded. Further details are available here: [https://cls-
44 genetics.github.io/docs/MCS.html](https://cls-genetics.github.io/docs/MCS.html)

45

46 PGS for both ALSPAC and MCS were calculated in individuals of genetically inferred
47 European ancestries. Genetic principal components were calculated for both cohorts
48 using PC-AiR as described earlier. We calculated PGS for $iPSYCH_{before11}$ and
49 $iPSYCH_{after10}$ and used these in all analyses in the MCS to keep it consistent with

1 analyses in SPARK where we could only use the iPSYCH GWAS to avoid overlap
2 between the training and testing sample.

3 *Association with SDQ*

4 We obtained scores on the SDQ total and subscales for six ages in the MCS and five
5 ages in ALSPAC. We ran cross-sectional analysis at each age using multiple linear
6 regression with PGS for iPSYCH_{before11} and iPSYCH_{after10}, with sex, age, and the first
7 10 genetic principal components as covariates. Additionally, we ran multiple linear
8 mixed effects regression using *lme4* package in R¹²², fitting a PGS by age interaction
9 term to investigate if the effects of PGS on SDQ change over time.

10
11
12 To investigate if the differences in association between MCS and ALSPAC were due
13 to differences in ascertainment between the two cohorts, we matched ALSPAC to
14 MCS using entropy balancing¹²³ and re-ran the PGS association analyses. Entropy
15 balancing is a reweighting technique that ensures the covariate distributions are
16 identical between groups. This method uses optimisation algorithms to assign weights
17 to individuals such that the weighted average of the covariates in ALSPAC (the larger
18 genotyped cohort) matches that of MCS (the smaller genotyped cohort), minimising
19 confounding biases and increasing comparability. We used the child's biological sex,
20 maternal age at delivery, and maternal highest educational qualification at first data
21 collection in each cohort as matching factors. We considered using propensity score
22 matching with a 1:1 ratio to obtain a well-balanced subsample of ALSPAC, yet this
23 approach would have resulted in substantial data loss and potential risk of residual
24 confounding due to limited covariate selection. Therefore, we opted for entropy
25 balancing to retain a larger sample size in the ALSPAC cohort. Entropy balancing was
26 conducted using the *ebal* package in R¹²⁴.

27 *Association with developmental milestones and autism diagnosis*

28 In ALSPAC, we obtained understanding of simple phrases (e.g., “do you want that”, or
29 “come here”) and gesture scores from the Macarthur-Bates Communicative
30 Development Inventories¹²⁵ at 15 months of age. We conducted multiple linear
31 regression using PGS for iPSYCH_{before11} and iPSYCH_{after10}, with sex, age, and the first
32 10 genetic principal components as covariates.

33
34
35 Autism diagnosis in the MCS was obtained using parent/caregiver reports of
36 autism/asperger syndrome diagnosis by a doctor at ages 5, 7, 11, and 14. We
37 identified individuals with an autism diagnosis at age 7 or earlier, age 11 or earlier, or
38 between ages 11 and 14. We conducted Firth's bias-reduced multiple logistic
39 regression (*logistf* package in R) using PGS for iPSYCH_{before11} and iPSYCH_{after10}, with
40 sex, age and the first 10 genetic principal components covariates.

41 42 43 **Code availability**

- 44 ● Lavaan (LGCM): <https://lavaan.ugent.be/tutorial/growth.html>
- 45 ● lcmm(GMM): <https://github.com/CecileProust-Lima/lcmm>
- 46 ● Softimpute: <https://cran.r-project.org/web/packages/softImpute/softImpute.pdf>
- 47 ● SPARK quality control, imputation and GWAS:
48 https://github.com/vwarrier/SPARK_iWES2_imputation/
- 49 ● Bespoke genetic analyses code:
50 https://github.com/vwarrier/autism_agediagnosis/

- 1 ● PRSCs: <https://github.com/getian107/PRSCs>
- 2 ● fastGWA and GCTA: <https://yanglab.westlake.edu.cn/software/gcta/#Overview>
- 3 ● GenomicSEM: <https://github.com/GenomicSEM/GenomicSEM>
- 4 ● LDSC: <https://github.com/bulik/ldsc>
- 5 ● KING: <https://www.kingrelatedness.com/manual.shtml>
- 6 ● Plink: <https://www.cog-genomics.org/plink/2.0/>
- 7 ● PC-AiR: <https://github.com/UW-GAC/GENESIS>
- 8 ● Lme4: <https://github.com/lme4/lme4/>
- 9 ● Logistf: <https://cran.r-project.org/web/packages/logistf/index.html>

12 Data availability:

- 13 ● SPARK autism GWAS: https://bitbucket.org/steinlabunc/spark_asd_sumstats/src
- 14 ● FinnGen autism GWAS: https://www.finnngen.fi/en/access_results
- 15 ● iPSYCH autism GWAS (unstratified, sex-stratified and age at diagnosis stratified) can be obtained from Anders Borglum and Jakob Grove.
- 16 ● Psychiatric GWAS summary stats: <https://pgc.unc.edu/>
- 17 ● GWAS educational attainment: <https://thessgac.com/>
- 18 ● GWAS cognitive aptitude: https://cnrc.nl/research/summary_statistics/
- 19 ● For ALSPAC, the study website contains details of all the data that is available through a fully searchable data dictionary and variable search tool": <http://www.bristol.ac.uk/alspac/researchers/our-data/>
- 20 ● For MCS, data can be obtain after application through the UK Data Service: <https://beta.ukdataservice.ac.uk/datacatalogue/series/series?id=2000031>
- 21 ● Summary statistics for the SPARK based GWAS will be made upon publication.

28 Acknowledgements

29 This research was supported by funding from the Simons Foundation for Autism
30 Research Initiative, the Wellcome Trust (214322\Z\18\Z), Horizon-Europe R2D2-MH
31 (grant agreement number 101057385), and UKRI (10063472). For the purpose of
32 open access, we have applied a CC BY public copyright licence to any author-
33 accepted manuscript version arising from this submission. S.B.-C. also received
34 funding from the Autism Centre of Excellence at Cambridge, the Templeton World
35 Charitable Fund, the MRC and the National Institute for Health Research Cambridge
36 Biomedical Research Centre. The research was supported by the National Institute for
37 Health Research Applied Research Collaboration East of England. Any views
38 expressed are those of the author(s) and not necessarily those of the funder. Some of
39 the results leading to this publication have received funding from the Innovative
40 Medicines Initiative 2 Joint Undertaking under grant agreement no. 777394 for the
41 project AIMS-2-TRIALS. This joint undertaking receives support from the European
42 Union's Horizon 2020 research and innovation program and the EFPIA and Autism
43 Speaks, Autistica and the SFARI. The iPSYCH team was supported by grants from
44 the Lundbeck Foundation (R102-A9118, R155-2014-1724 and R248-2017-2003), the
45 NIMH (1R01MH124851-01 to A.D.B.), and EU's Horizon Europe program (R2D2-MH;
46 grant agreement no. 101057385 to A.D.B.). The Danish National Biobank resource
47 was supported by the Novo Nordisk Foundation. High-performance computer capacity
48 for handling and statistical analysis of iPSYCH data on the GenomeDK HPC facility
49 was provided by the Center for Genomics and Personalized Medicine and the Centre

1 for Integrative Sequencing, iSEQ, Aarhus University, Denmark (grant to A.D.B.). The
2 UK Medical Research Council and Wellcome (Grant ref: 217065/Z/19/Z) and the
3 University of Bristol provide core support for ALSPAC. This publication is the work of
4 the authors and the authors will serve as guarantors for the contents of this paper. A
5 comprehensive list of grants funding is available on the ALSPAC website
6 (<http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf>).
7 R2D2-MH has been funded by Horizon Europe [grant agreement no. 101057385], by
8 UK Research and Innovation (UKRI) under the UK government's Horizon Europe
9 funding guarantee [grant no.10039383] and by the Swiss State Secretariat for
10 Education, Research and Innovation (SERI) under contract number 22.00277

11

12 We are grateful to the Centre for Longitudinal Studies (CLS), UCL Social Research
13 Institute, for the use of these data and to the UK Data Service for making them
14 available. However, neither CLS nor the UK Data Service bear any responsibility for
15 the analysis or interpretation of these data. This paper uses unit record data from
16 Growing Up in Australia, the Longitudinal Study of Australian Children. The study is
17 conducted in partnership between the Department of Social Services (DSS), the
18 Australian Institute of Family Studies (AIFS) and the Australian Bureau of Statistics
19 (ABS). The findings and views reported in this paper are those of the author and should
20 not be attributed to DSS, AIFS or the ABS. Growing Up in Ireland (GUI) has been
21 funded by the Government of Ireland through the Department of Children, Equality,
22 Disability, Integration and Youth (DCEDIY) and the Central Statistics Office (CSO).
23 Results in this report are based on analysis of data from Research Microdata Files
24 provided by the Central Statistics Office (CSO). Neither the CSO nor the DCEDIY take
25 any responsibility for the views expressed or the outputs generated from these
26 analyses. We are extremely grateful to all the families who took part in this study, the
27 midwives for their help in recruiting them, and the whole ALSPAC team, which includes
28 interviewers, computer and laboratory technicians, clerical workers, research
29 scientists, volunteers, managers, receptionists and nurses.

30 We thank Alex Kwong, Tamsin Ford, Will Mandy, and Andrew Grotzinger for helpful
31 discussions.

32 **Ethics declarations**

33 ADB received speakers' fee from Lundbeck. The authors declare no competing
34 interests.

35

36 **Author contributions**

37 XZ and VW conducted most of the analyses, with the remainder being conducted by JG, YG,
38 CKB, and LKN. MK, EV, AG, LH, AH, AR, BSt.P, and ADB provided summary statistics for
39 various analyses. SASN, DSM, and EMW carried out data preparation and quality control,
40 with assistance from KES, VKC, PD, SL, TM, MK, SA and DB. VW supervised the analyses
41 and directed the study with inputs from HCM and JG. VW and XZ wrote the initial draft with
42 input from HCM. DSM, EW, DHG, NRW, EBR, TB, and SBC provided intellectual input. All
43 authors read and commented on the final manuscript.

44

1 **APEX Consortium**

2 Deep Adhya, Carrie Allison, Bonnie Ayeung, Rosie Bamford, Simon Baron-Cohen,
3 Richard Bethlehem, Tal Biron-Shental, Graham Burton, Wendy Cowell, Jonathan
4 Davies, Joanna Davis, Dori Floris, Alice Franklin, Lidia Gabis, Daniel Geschwind,
5 Ramin Ali Marandi Ghoddousi, David M. Greenberg, Yuanjun Gu, Alexandra Havdahl,
6 Alexander Heazell, Rosemary Holt, Matthew Hurles, Yumnah Khan, Meng-Chuan Lai,
7 Madeline Lancaster, Michael Lombardo, Hilary Martin, Jose Gonzalez Martinez,
8 Jonathan Mill, Mahmoud Musa, Kathy Niakan, Adam Pavlinek, Lucia Dutan Polit,
9 Marcin Radecki, David Rowitch, Jenifer Sakai, Laura Sichlinger, Deepak Srivastava,
10 Alexandros Tsompanidis, Florina Uzefovsky, Varun Warriar, Elizabeth Weir, Xinhe
11 Zhang.

12

13 **iPSYCH Autism working group**

14 Anders Borglum, Jonas Bybjerg-Grauholm, Jakob Grove, David M. Hougaard, Ole
15 Mors, Preben Bo Mortensen, Merete Nordentoft and Thomas Werge.

16

17 **PGC-PTSD consortium**

18 Caroline M. Nievergelt, Adam X. Maihofer, Elizabeth G. Atkinson, Chia-Yen Chen, Karmel W.
19 Choi, Jonathan R. I. Coleman, Nikolaos P. Daskalakis, Laramie E. Duncan, Renato Polimanti,
20 Cindy Aaronson, Ananda B. Amstadter, Soren B. Andersen, Ole A. Andreassen, Paul A.
21 Arbisi, Allison E. Ashley-Koch, S. Bryn Austin, Esmina Avdibegović, Dragan Babić, Silviu-Alin
22 Bacanu, Dewleen G. Baker, Anthony Batzler, Jean C. Beckham, Sintia Belangero, Corina
23 Benjet, Carisa Bergner, Linda M. Bierer, Joanna M. Biernacka, Laura J. Bierut, Jonathan I.
24 Bisson, Marco P. Boks, Elizabeth A. Bolger, Amber Brandolino, Gerome Breen, Rodrigo
25 Affonseca Bressan, Richard A. Bryant, Angela C. Bustamante, Jonas Bybjerg-Grauholm,
26 Marie Bækvad-Hansen, Anders D. Børglum, Sigrid Børte, Leah Cahn, Joseph R. Calabrese,
27 Jose Miguel Caldas-de-Almeida, Chris Chatzinakos, Sheraz Cheema, Sean A. P. Clouston,
28 Lucía Colodro-Conde, Brandon J. Coombes, Carlos S. Cruz-Fuentes, Anders M. Dale,
29 Shareefa Dalvie, Lea K. Davis, Jürgen Deckert, Douglas L. Delahanty, Michelle F. Dennis,
30 Frank Desarnaud, Christopher P. DiPietro, Seth G. Disner, Anna R. Docherty, Katharina
31 Domschke, Grete Dyb, Alma Džubur Kulenović, Howard J. Edenberg, Alexandra Evans,
32 Chiara Fabbri, Negar Fani, Lindsay A. Farrer, Adriana Feder, Norah C. Feeny, Janine D. Flory,
33 David Forbes, Carol E. Franz, Sandro Galea, Melanie E. Garrett, Bizu Gelaye, Joel Gelernter,
34 Elbert Geuze, Charles F. Gillespie, Slavina B. Goleva, Scott D. Gordon, Aferdita Goçi, Lana
35 Ruvolo Grasser, Camila Guindalini, Magali Haas, Saskia Hagenaars, Michael A. Hauser,
36 Andrew C. Heath, Sian M. J. Hemmings, Victor Hesselbrock, Ian B. Hickie, Kelleigh Hogan,
37 David Michael Hougaard, Hailiang Huang, Laura M. Huckins, Kristian Hveem, Miro
38 Jakovljević, Arash Javanbakht, Gregory D. Jenkins, Jessica Johnson, Ian Jones, Tanja
39 Jovanovic, Karen-Inge Karstoft, Milissa L. Kaufman, James L. Kennedy, Ronald C. Kessler,
40 Alaptagin Khan, Nathan A. Kimbrel, Anthony P. King, Nastassja Koen, Roman Kotov, Henry
41 R. Kranzler, Kristi Krebs, William S. Kremen, Pei-Fen Kuan, Bruce R. Lawford, Lauren A. M.
42 Lebois, Kelli Lehto, Daniel F. Levey, Catrin Lewis, Israel Liberzon, Sarah D. Linnstaedt, Mark
43 W. Logue, Adriana Lori, Yi Lu, Benjamin J. Luft, Michelle K. Lupton, Jurjen J. Luykx, Iouri
44 Makotkine, Jessica L. Maples-Keller, Shelby Marchese, Charles Marmar, Nicholas G. Martin,

1 Gabriela A. Martínez-Levy, Kerrie McAloney, Alexander McFarlane, Katie A. McLaughlin,
2 Samuel A. McLean, Sarah E. Medland, Divya Mehta, Jacquelyn Meyers, Vasiliki Michopoulos,
3 Elizabeth A. Mikita, Lili Milani, William Milberg, Mark W. Miller, Rajendra A. Morey, Charles
4 Phillip Morris, Ole Mors, Preben Bo Mortensen, Mary S. Mufford, Elliot C. Nelson, Merete
5 Nordentoft, Sonya B. Norman, Nicole R. Nugent, Meaghan O'Donnell, Holly K. Orcutt, Pedro
6 M. Pan, Matthew S. Panizzon, Gita A. Pathak, Edward S. Peters, Alan L. Peterson, Matthew
7 Peverill, Robert H. Pietrzak, Melissa A. Polusny, Bernice Porjesz, Abigail Powers, Xue-Jun
8 Qin, Andrew Ratanatharathorn, Victoria B. Risbrough, Andrea L. Roberts, Alex O. Rothbaum,
9 Barbara O. Rothbaum, Peter Roy-Byrne, Kenneth J. Ruggiero, Ariane Rung, Heiko Runz, Bart
10 P. F. Rutten, Stacey Saenz de Viteri, Giovanni Abrahão Salum, Laura Sampson, Sixto E.
11 Sanchez, Marcos Santoro, Carina Seah, Soraya Seedat, Julia S. Seng, Andrey Shabalin,
12 Christina M. Sheerin, Derrick Silove, Alicia K. Smith, Jordan W. Smoller, Scott R. Sponheim,
13 Dan J. Stein, Synne Stensland, Jennifer S. Stevens, Jennifer A. Sumner, Martin H. Teicher,
14 Wesley K. Thompson, Arun K. Tiwari, Edward Trapido, Monica Uddin, Robert J. Ursano,
15 Unnur Valdimarsdóttir, Miranda Van Hooff, Eric Vermetten, Christiaan H. Vinkers, Joanne
16 Voisey, Yunpeng Wang, Zhewu Wang, Monika Waszczuk, Heike Weber, Frank R. Wendt,
17 Thomas Werge, Michelle A. Williams, Douglas E. Williamson, Bendik S. Winsvold, Sherry
18 Winternitz, Christiane Wolf, Erika J. Wolf, Yan Xia, Ying Xiong, Rachel Yehuda, Keith A.
19 Young, Ross McD Young, Clement C. Zai, Gwyneth C. Zai, Mark Zervas, Hongyu Zhao, Lori
20 A. Zoellner, John-Anker Zwart, Terri deRoon-Cassini, Sanne J. H. van Rooij, Leigh L. van den
21 Heuvel, AURORA Study, Estonian Biobank Research Team, FinnGen Investigators, HUNT
22 All-In Psychiatry, Murray B. Stein, Kerry J. Ressler and Karestan C. Koenen

23

24

25

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28

References

1. Lord, C. *et al.* Autism spectrum disorder. *Nat Rev Dis Primers* **6**, 5 (2020).
2. Kanner, L. Autistic disturbances of affective contact. *Nervous Child: Journal of Psychopathology, Psychotherapy, Mental Hygiene, and Guidance of the Child* **2** 217–50 (1943).
3. Asperger, H. ‘Autistic psychopathy’ in childhood. in *Autism and Asperger syndrome* (ed. Frith, U.) 37–92 (Cambridge University Press, Cambridge, 1944).
4. Schendel, D. E. & Thorsteinsson, E. Cumulative Incidence of Autism Into Adulthood for Birth Cohorts in Denmark, 1980-2012. *JAMA* **320**, 1811–1813 (2018).
5. Russell, G. *et al.* Time trends in autism diagnosis over 20 years: a UK population-based cohort study. *J. Child Psychol. Psychiatry* **63**, 674–682 (2022).
6. Jensen, C. M., Steinhausen, H.-C. & Lauritsen, M. B. Time trends over 16 years in incidence-rates of autism spectrum disorders across the lifespan based on nationwide Danish register data. *J. Autism Dev. Disord.* **44**, 1808–1818 (2014).
7. Ozonoff, S. *et al.* Diagnosis of Autism Spectrum Disorder After Age 5 in Children Evaluated Longitudinally Since Infancy. *J. Am. Acad. Child Adolesc. Psychiatry* **57**, 849–857.e2 (2018).
8. Mandy, W. *et al.* Mental health and social difficulties of late-diagnosed autistic children, across childhood and adolescence. *J. Child Psychol. Psychiatry* **63**, 1405–1414 (2022).
9. May, T., Brignell, A. & Williams, K. Parent-reported autism diagnostic stability and trajectories in the Longitudinal Study of Australian Children. *Autism Res.* **14**, 773–786 (2021).
10. Allison, C. *et al.* Quantitative Checklist for Autism in Toddlers (Q-CHAT). A population screening study with follow-up: the case for multiple time-point screening for autism. *BMJ Paediatr Open* **5**, e000700 (2021).
11. Davidovitch, M., Levit-Binnun, N., Golan, D. & Manning-Courtney, P. Late diagnosis of autism spectrum disorder after initial negative assessment by a multidisciplinary team.

- 1 *J. Dev. Behav. Pediatr.* **36**, 227–234 (2015).
- 2 12. Avlund, S. H. *et al.* Factors Associated with a Delayed Autism Spectrum Disorder
3 Diagnosis in Children Previously Assessed on Suspicion of Autism. *J. Autism Dev.*
4 *Disord.* **51**, 3843–3856 (2021).
- 5 13. Bazelmans, T. *et al.* Mid-childhood autism sibling recurrence in infants with a family
6 history of autism. *Autism Res.* (2024) doi:10.1002/aur.3182.
- 7 14. Landa, R. J., Reetzke, R., Holingue, C. B., Herman, D. & Hess, C. R. Diagnostic
8 Stability and Phenotypic Differences Among School-Age Children Diagnosed With ASD
9 Before Age 2. *Front. Psychiatry* **13**, 805686 (2022).
- 10 15. Austerberry, C., Mateen, M., Fearon, P. & Ronald, A. Heritability of Psychological Traits
11 and Developmental Milestones in Infancy: A Systematic Review and Meta-analysis.
12 *JAMA Netw Open* **5**, e2227887 (2022).
- 13 16. Hegemann, L. *et al.* Genetic and phenotypic heterogeneity in early neurodevelopmental
14 traits in the Norwegian Mother, Father and Child Cohort Study. *Mol. Autism* **15**, 25
15 (2024).
- 16 17. St Pourcain, B. *et al.* Variability in the common genetic architecture of social-
17 communication spectrum phenotypes during childhood and adolescence. *Mol. Autism* **5**,
18 18 (2014).
- 19 18. Rødgaard, E.-M., Jensen, K., Miskowiak, K. W. & Mottron, L. Autism comorbidities show
20 elevated female-to-male odds ratios and are associated with the age of first autism
21 diagnosis. *Acta Psychiatr. Scand.* **144**, 475–486 (2021).
- 22 19. Jadav, N. & Bal, V. H. Associations between co-occurring conditions and age of autism
23 diagnosis: Implications for mental health training and adult autism research. *Autism*
24 *Res.* **15**, 2112–2125 (2022).
- 25 20. Goodman, R. Strengths and Difficulties Questionnaire. *Child and Adolescent Psychiatry*
26 *and Mental Health European Journal of Psychological Assessment Psychological*
27 *Assessment School Psychology Quarterly Clinical Psychologist* doi:10.1037/t00540-000.
- 28 21. Stone, L. L. *et al.* The Strengths and Difficulties Questionnaire: psychometric properties

- 1 of the parent and teacher version in children aged 4-7. *BMC Psychol* **3**, 4 (2015).
- 2 22. Kovacs, S. & Sharp, C. Criterion validity of the Strengths and Difficulties Questionnaire
3 (SDQ) with inpatient adolescents. *Psychiatry Res.* **219**, 651–657 (2014).
- 4 23. Borg, A.-M., Kaukonen, P., Salmelin, R., Joukamaa, M. & Tamminen, T. Reliability of
5 the strengths and difficulties questionnaire among Finnish 4-9-year-old children. *Nord. J.*
6 *Psychiatry* **66**, 403–413 (2012).
- 7 24. Speyer, L. G., Auyeung, B. & Murray, A. L. Longitudinal Invariance of the Strengths and
8 Difficulties Questionnaire Across Ages 4 to 16 in the ALSPAC Sample. *Assessment* **30**,
9 1884–1894 (2023).
- 10 25. Murray, A. L., Speyer, L. G., Hall, H. A., Valdebenito, S. & Hughes, C. A Longitudinal
11 and Gender Invariance Analysis of the Strengths and Difficulties Questionnaire Across
12 Ages 3, 5, 7, 11, 14, and 17 in a Large U.K.-Representative Sample. *Assessment* **29**,
13 1248–1261 (2022).
- 14 26. Woerner, W. *et al.* The Strengths and Difficulties Questionnaire overseas: Evaluations
15 and applications of the SDQ beyond Europe. *Eur. Child Adolesc. Psychiatry* **13**, ii47–
16 ii54 (2004).
- 17 27. Dalsgaard, S. *et al.* Incidence Rates and Cumulative Incidences of the Full Spectrum of
18 Diagnosed Mental Disorders in Childhood and Adolescence. *JAMA Psychiatry* **77**, 155–
19 164 (2020).
- 20 28. Dahl, R. E., Allen, N. B., Wilbrecht, L. & Suleiman, A. B. Importance of investing in
21 adolescence from a developmental science perspective. *Nature* **554**, 441–450 (2018).
- 22 29. Foulkes, L. & Blakemore, S.-J. Studying individual differences in human adolescent
23 brain development. *Nat. Neurosci.* **21**, 315–323 (2018).
- 24 30. Angold, A., Costello, E., Messer, S. & Pickles, A. Development of a short questionnaire
25 for use in epidemiological studies of depression in children and adolescents: Factor
26 composition and structure across development. *Int. J. Methods Psychiatr. Res.* **5**, 251–
27 262 (1995).
- 28 31. Hrdlicka, M. *et al.* Predictors of age at diagnosis in autism spectrum disorders: the use

- 1 of multiple regression analyses and a classification tree on a clinical sample. *Eur. Child*
2 *Adolesc. Psychiatry* **33**, 1171–1177 (2024).
- 3 32. Brett, D., Warnell, F., McConachie, H. & Parr, J. R. Factors Affecting Age at ASD
4 Diagnosis in UK: No Evidence that Diagnosis Age has Decreased Between 2004 and
5 2014. *Journal of Autism and Developmental Disorders* vol. 46 1974–1984 Preprint at
6 <https://doi.org/10.1007/s10803-016-2716-6> (2016).
- 7 33. St Pourcain, B. *et al.* Heritability and genome-wide analyses of problematic peer
8 relationships during childhood and adolescence. *Hum. Genet.* **134**, 539–551 (2015).
- 9 34. Knafo, A. & Plomin, R. Prosocial behavior from early to middle childhood: genetic and
10 environmental influences on stability and change. *Dev. Psychol.* **42**, 771–786 (2006).
- 11 35. Stergiakouli, E. *et al.* Shared genetic influences between dimensional ASD and ADHD
12 symptoms during child and adolescent development. *Mol. Autism* **8**, 18 (2017).
- 13 36. Jami, E. S. *et al.* Genome-wide Association Meta-analysis of Childhood and Adolescent
14 Internalizing Symptoms. *J. Am. Acad. Child Adolesc. Psychiatry* **61**, 934–945 (2022).
- 15 37. St Pourcain, B. *et al.* ASD and schizophrenia show distinct developmental profiles in
16 common genetic overlap with population-based social communication difficulties. *Mol.*
17 *Psychiatry* (2017) doi:10.1038/mp.2016.198.
- 18 38. St Pourcain, B. *et al.* Developmental Changes Within the Genetic Architecture of Social
19 Communication Behavior: A Multivariate Study of Genetic Variance in Unrelated
20 Individuals. *Biol. Psychiatry* **83**, 598–606 (2018).
- 21 39. SPARK Consortium. Electronic address: pfeliciano@simonsfoundation.org & SPARK
22 Consortium. SPARK: A US Cohort of 50,000 Families to Accelerate Autism Research.
23 *Neuron* **97**, 488–493 (2018).
- 24 40. Demontis, D. *et al.* Genome-wide analyses of ADHD identify 27 risk loci, refine the
25 genetic architecture and implicate several cognitive domains. *Nat. Genet.* **55**, 198–208
26 (2023).
- 27 41. Trubetskoy, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in
28 schizophrenia. *Nature* **604**, 502–508 (2022).

- 1 42. Fu, J. M. *et al.* Rare coding variation provides insight into the genetic architecture and
2 phenotypic context of autism. *Nat. Genet.* **54**, 1320–1331 (2022).
- 3 43. Satterstrom, F. K. *et al.* Large-Scale Exome Sequencing Study Implicates Both
4 Developmental and Functional Changes in the Neurobiology of Autism. *Cell* **180**, 568–
5 584.e23 (2020).
- 6 44. Als, T. D. *et al.* Depression pathophysiology, risk prediction of recurrence and comorbid
7 psychiatric disorders using genome-wide analyses. *Nat. Med.* **29**, 1832–1844 (2023).
- 8 45. Lee, J. J. *et al.* Gene discovery and polygenic prediction from a genome-wide
9 association study of educational attainment in 1.1 million individuals. *Nat. Genet.* **50**,
10 1112–1121 (2018).
- 11 46. Savage, J. E. *et al.* Genome-wide association meta-analysis in 269,867 individuals
12 identifies new genetic and functional links to intelligence. *Nat. Genet.* **1** (2018).
- 13 47. Sada-Fuente, E. *et al.* Common genetic variants contribute to heritability of age at onset
14 of schizophrenia. *Transl. Psychiatry* **13**, 201 (2023).
- 15 48. Harder, A. *et al.* Genetics of age-at-onset in major depression. *Transl. Psychiatry* **12**,
16 124 (2022).
- 17 49. Feng, Y.-C. A. *et al.* Findings and insights from the genetic investigation of age of first
18 reported occurrence for complex disorders in the UK Biobank and FinnGen. *bioRxiv*
19 (2020) doi:10.1101/2020.11.20.20234302.
- 20 50. Warriar, V. *et al.* Genetic correlates of phenotypic heterogeneity in autism. *Nat. Genet.*
21 (2022) doi:10.1038/s41588-022-01072-5.
- 22 51. The Autism Spectrum Disorders Working Group of The Psychiatric Genomics
23 Consortium. Meta-analysis of GWAS of over 16,000 individuals with autism spectrum
24 disorder highlights a novel locus at 10q24.32 and a significant overlap with
25 schizophrenia. *Mol. Autism* **8**, 21 (2017).
- 26 52. Matoba, N. *et al.* Common genetic risk variants identified in the SPARK cohort support
27 DDHD2 as a candidate risk gene for autism. *Transl. Psychiatry* **10**, 265 (2020).
- 28 53. Patalay, P. & Gage, S. H. Changes in millennial adolescent mental health and health-

- 1 related behaviours over 10 years: a population cohort comparison study. *Int. J.*
2 *Epidemiol.* **48**, 1650–1664 (2019).
- 3 54. Grotzinger, A. D. *et al.* Genomic structural equation modelling provides insights into the
4 multivariate genetic architecture of complex traits. *Nature Human Behaviour* **1** (2019).
- 5 55. Dawson, G., Rieder, A. D. & Johnson, M. H. Prediction of autism in infants: progress
6 and challenges. *Lancet Neurol.* **22**, 244–254 (2023).
- 7 56. Wray, N. R. *et al.* Genome-wide association analyses identify 44 risk variants and refine
8 the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018).
- 9 57. Sheldrick, R. C., Maye, M. P. & Carter, A. S. Age at First Identification of Autism
10 Spectrum Disorder: An Analysis of Two US Surveys. *J. Am. Acad. Child Adolesc.*
11 *Psychiatry* **56**, 313–320 (2017).
- 12 58. Bulik-Sullivan, B. K. *et al.* An atlas of genetic correlations across human diseases and
13 traits. *Nat. Genet.* **47**, 1236–1241 (2015).
- 14 59. Cross-Disorder Group of the Psychiatric Genomics Consortium *et al.* Genetic
15 relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat.*
16 *Genet.* **45**, 984–994 (2013).
- 17 60. Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum
18 disorder. *Nat. Genet.* **51**, 431–444 (2019).
- 19 61. Rajagopal, V. M. *et al.* Differences in the genetic architecture of common and rare
20 variants in childhood, persistent and late-diagnosed attention-deficit hyperactivity
21 disorder. *Nat. Genet.* **54**, 1117–1124 (2022).
- 22 62. Breunig, S. *et al.* Examining Differences in the Genetic and Functional Architecture of
23 Attention-Deficit/Hyperactivity Disorder Diagnosed in Childhood and Adulthood. *Biol*
24 *Psychiatry Glob Open Sci* **4**, 100307 (2024).
- 25 63. Martini, M. I. *et al.* Sex Differences in Mental Health Problems and Psychiatric
26 Hospitalization in Autistic Young Adults. *JAMA Psychiatry* **79**, 1188–1198 (2022).
- 27 64. Dworzynski, K., Ronald, A., Bolton, P. & Happé, F. How different are girls and boys
28 above and below the diagnostic threshold for autism spectrum disorders? *J. Am. Acad.*

- 1 *Child Adolesc. Psychiatry* **51**, 788–797 (2012).
- 2 65. Milner, V. *et al.* Does camouflaging predict age at autism diagnosis? A comparison of
3 autistic men and women. *Autism Res.* **17**, 626–636 (2024).
- 4 66. Dean, M., Harwood, R. & Kasari, C. The art of camouflage: Gender differences in the
5 social behaviors of girls and boys with autism spectrum disorder. *Autism* **21**, 678–689
6 (2017).
- 7 67. Hull, L. *et al.* Development and Validation of the Camouflaging Autistic Traits
8 Questionnaire (CAT-Q). *J. Autism Dev. Disord.* **49**, 819–833 (2019).
- 9 68. Cook, J., Crane, L., Hull, L., Bourne, L. & Mandy, W. Self-reported camouflaging
10 behaviours used by autistic adults during everyday social interactions. *Autism* **26**, 406–
11 421 (2022).
- 12 69. Milner, V., Mandy, W., Happé, F. & Colvert, E. Sex differences in predictors and
13 outcomes of camouflaging: Comparing diagnosed autistic, high autistic trait and low
14 autistic trait young adults. *Autism* **27**, 402–414 (2023).
- 15 70. Ross, A., Grove, R. & McAloon, J. The relationship between camouflaging and mental
16 health in autistic children and adolescents. *Autism Res.* **16**, 190–199 (2023).
- 17 71. Perry, E., Mandy, W., Hull, L. & Cage, E. Understanding Camouflaging as a Response
18 to Autism-Related Stigma: A Social Identity Theory Approach. *J. Autism Dev. Disord.*
19 **52**, 800–810 (2022).
- 20 72. Roman-Urrestarazu, A., Dumas, G. & Warrier, V. Naming Autism in the Right Context.
21 *JAMA Pediatr.* **176**, 633–634 (2022).
- 22 73. Connelly, R. & Platt, L. Cohort profile: UK Millennium Cohort Study (MCS). *Int. J.*
23 *Epidemiol.* **43**, 1719–1725 (2014).
- 24 74. Layte, R. & McCrory, C. Growing Up in Ireland: Maternal Health Behaviours and Child
25 Growth in Infancy. *Research Series* (2015).
- 26 75. Clifford, S. A., Davies, S., Wake, M. & Child Health CheckPoint Team. Child Health
27 CheckPoint: cohort summary and methodology of a physical health and biospecimen
28 module for the Longitudinal Study of Australian Children. *BMJ Open* **9**, 3–22 (2019).

- 1 76. Anderson, E. R. Analyzing change in short-term longitudinal research using cohort-
2 sequential designs. *J. Consult. Clin. Psychol.* **61**, 929–940 (1993).
- 3 77. Russell, G., Rodgers, L. R. & Ford, T. The strengths and difficulties questionnaire as a
4 predictor of parent-reported diagnosis of autism spectrum disorder and attention deficit
5 hyperactivity disorder. *PLoS One* **8**, e80247 (2013).
- 6 78. Grasso, M., Lazzaro, G., Demaria, F., Menghini, D. & Vicari, S. The Strengths and
7 Difficulties Questionnaire as a Valuable Screening Tool for Identifying Core Symptoms
8 and Behavioural and Emotional Problems in Children with Neuropsychiatric Disorders.
9 *Int. J. Environ. Res. Public Health* **19**, (2022).
- 10 79. Roman-Urrestarazu, A. *et al.* Association of Race/Ethnicity and Social Disadvantage
11 With Autism Prevalence in 7 Million School Children in England. *JAMA Pediatr.* **175**,
12 e210054 (2021).
- 13 80. Terhaag, S., Fitzsimons, E., Daraganova, G. & Patalay, P. Sex, ethnic and
14 socioeconomic inequalities and trajectories in child and adolescent mental health in
15 Australia and the UK: findings from national prospective longitudinal studies. *J. Child*
16 *Psychol. Psychiatry* **62**, 1255–1267 (2021).
- 17 81. Pender, R., Fearon, P., St Pourcain, B., Heron, J. & Mandy, W. Developmental
18 trajectories of autistic social traits in the general population. *Psychol. Med.* 1–9 (2021).
- 19 82. Rosseel, Y. Lavaan: An R package for structural equation modeling and more. Version
20 0.5--12 (BETA). *J. Stat. Softw.* **48**, 1–36 (2012).
- 21 83. Hoeksma, J. B. & Kelderman, H. On growth curves and mixture models. *Infant Child*
22 *Dev.* **15**, 627–634 (2006).
- 23 84. Proust-Lima, C., Philipps, V. & Liqueur, B. Estimation of Extended Mixed Models Using
24 Latent Classes and Latent Processes: The R Package lcmm. *J. Stat. Softw.* **78**, 1–56
25 (2017).
- 26 85. Ram, N. & Grimm, K. J. Growth Mixture Modeling: A Method for Identifying Differences
27 in Longitudinal Change Among Unobserved Groups. *Int. J. Behav. Dev.* **33**, 565–576
28 (2009).

- 1 86. Grimm, K. J., Mazza, G. L. & Davoudzadeh, P. Model Selection in Finite Mixture
2 Models: A k-Fold Cross-Validation Approach. *Struct. Equ. Modeling* **24**, 246–256
3 (2017).
- 4 87. Tonidandel, S. & LeBreton, J. M. Relative Importance Analysis: A Useful Supplement to
5 Regression Analysis. *J. Bus. Psychol.* **26**, 1–9 (2011).
- 6 88. Leaf, J. B. *et al.* Increasing social skills and pro-social behavior for three children
7 diagnosed with autism through the use of a teaching package. *Res. Autism Spectr.*
8 *Disord.* **3**, 275–289 (2009).
- 9 89. Azen, R. & Budescu, D. V. Comparing Predictors in Multivariate Regression Models: An
10 Extension of Dominance Analysis. *J. Educ. Behav. Stat.* **31**, 157–180 (2006).
- 11 90. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies.
12 *Bioinformatics* **26**, 2867–2873 (2010).
- 13 91. Conomos, M. P., Miller, M. B. & Thornton, T. A. Robust inference of population structure
14 for ancestry prediction and correction of stratification in the presence of relatedness.
15 *Genet. Epidemiol.* **39**, 276–293 (2015).
- 16 92. Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed
17 Program. *Nature* **590**, 290–299 (2021).
- 18 93. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.*
19 **48**, 1284–1287 (2016).
- 20 94. Loh, P.-R., Palamara, P. F. & Price, A. L. Fast and accurate long-range phasing in a UK
21 Biobank cohort. *Nat. Genet.* **48**, 811–816 (2016).
- 22 95. Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C. A. & Smoller, J. W. Polygenic prediction via
23 Bayesian regression and continuous shrinkage priors. *Nat. Commun.* **10**, 1776 (2019).
- 24 96. Bybjerg-Grauholm, J. *et al.* The iPSYCH2015 Case-Cohort sample: updated directions
25 for unravelling genetic and environmental architectures of severe mental disorders.
26 Preprint at <https://doi.org/10.1101/2020.11.30.20237768>.
- 27 97. Mullins, N. *et al.* Genome-wide association study of more than 40,000 bipolar disorder
28 cases provides new insights into the underlying biology. *Nat. Genet.* **53**, 817–829

- 1 (2021).
- 2 98. Okbay, A. *et al.* Polygenic prediction of educational attainment within and between
3 families from genome-wide association analyses in 3 million individuals. *Nat. Genet.* **54**,
4 437–449 (2022).
- 5 99. Dsm-iv. *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition*.
6 (American Psychiatric Publishing, Inc., Washington, DC, 1994).
- 7 100. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based
8 linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
- 9 101. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and
10 richer datasets. *Gigascience* **4**, 7 (2015).
- 11 102. Demange, P. A. *et al.* Estimating effects of parents' cognitive and non-cognitive skills on
12 offspring education using polygenic scores. *Nat. Commun.* **13**, 4801 (2022).
- 13 103. Weiner, D. J. *et al.* Polygenic transmission disequilibrium confirms that common and
14 rare variation act additively to create risk for autism spectrum disorders. *Nat. Genet.* **49**,
15 978–985 (2017).
- 16 104. Koko, M. E. *et al.* Contribution of autosomal rare and de novo variants to sex differences
17 in autism. *medRxiv* 2024–2004 (2024).
- 18 105. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in
19 141,456 humans. *Nature* **581**, 434–443 (2020).
- 20 106. Samocha, K. E. *et al.* Regional missense constraint improves variant deleteriousness
21 prediction. *bioRxiv* 148353 (2017) doi:10.1101/148353.
- 22 107. Jiang, L. *et al.* A resource-efficient tool for mixed model association analysis of large-
23 scale data. *Nat. Genet.* **51**, 1749–1755 (2019).
- 24 108. Lam, M. *et al.* RICOPILI: Rapid Imputation for COnsortias PIpeLine. *Bioinformatics* **36**,
25 930–933 (2020).
- 26 109. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and
27 accurate genotype imputation in genome-wide association studies through pre-phasing.
28 *Nat. Genet.* **44**, 955–959 (2012).

- 1 110. Iglesias, A. I. *et al.* Haplotype reference consortium panel: Practical implications of
2 imputations with large reference panels. *Hum. Mutat.* **38**, 1025–1032 (2017).
- 3 111. Jiang, L., Zheng, Z., Fang, H. & Yang, J. A generalized linear mixed model association
4 tool for biobank-scale data. *Nat. Genet.* **53**, 1616–1621 (2021).
- 5 112. Peyrot, W. J., Boomsma, D. I., Penninx, B. W. J. H. & Wray, N. R. Disease and
6 polygenic architecture: Avoid trio design and appropriately account for unscreened
7 control subjects for common disease. *Am. J. Hum. Genet.* **98**, 382–391 (2016).
- 8 113. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide
9 complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
- 10 114. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human
11 height. *Nat. Genet.* **42**, 565–569 (2010).
- 12 115. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from
13 polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
- 14 116. Kurki, M. I. *et al.* FinnGen provides genetic insights from a well-phenotyped isolated
15 population. *Nature* **613**, 508–518 (2023).
- 16 117. Golding, J., Pembrey, M., Jones, R. & ALSPAC Study Team. ALSPAC--the Avon
17 Longitudinal Study of Parents and Children. I. Study methodology. *Paediatr. Perinat.*
18 *Epidemiol.* **15**, 74–87 (2001).
- 19 118. Boyd, A. *et al.* Cohort Profile: The 'Children of the 90s'—the index offspring of the Avon
20 Longitudinal Study of Parents and Children. *Int. J. Epidemiol.* **42**, 111–127 (2013).
- 21 119. Fraser, A. *et al.* Cohort Profile: the Avon Longitudinal Study of Parents and Children:
22 ALSPAC mothers cohort. *Int. J. Epidemiol.* **42**, 97–110 (2013).
- 23 120. Fitzsimons, E. *et al.* Collection of genetic data at scale for a nationally representative
24 population: the UK Millennium Cohort Study. *Longit. Life Course Stud.* **13**, 169–187
25 (2021).
- 26 121. Pain, O. *et al.* Evaluation of polygenic prediction methodology within a reference-
27 standardized framework. *PLoS Genet.* **17**, e1009021 (2021).
- 28 122. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models

- 1 Using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).
- 2 123. Hainmueller, J. Entropy Balancing for Causal Effects: A Multivariate Reweighting
3 Method to Produce Balanced Samples in Observational Studies. *Polit. Anal.* **20**, 25–46
4 (2012).
- 5 124. ebal package - RDocumentation.
6 <https://rdocumentation.org/packages/ebal/versions/0.1-8>.
- 7 125. Fenson, L. *et al.* MacArthur-Bates Communicative Development Inventories, Second
8 Edition. doi:10.1037/t11538-000.