Polygenic scores for intelligence, educational attainment and schizophrenia are differentially associated with core autism features, IQ, and adaptive behaviour in autistic individuals

Varun Warrier, Claire Leblond, Freddy Cliquet, Aims-2-Trials Leap, Thomas Bourgeron, Simon Baron-Cohen

Polygenic scores for intelligence, educational attainment and schizophrenia are differentially associated with core autism features, IQ, and adaptive behaviour in autistic individuals

Varun Warrier[1,3], Claire S Leblond[2], Freddy Cliquet[2], AIMS-2-TRIALS LEAP, Thomas Bourgeron[2,4], and Simon Baron-Cohen[1,3,4]

1. Autism Research Centre, Department of Psychiatry, University of Cambridge

2. Human Genetics and Cognitive Functions, Institut Pasteur, UMR3571 CNRS, Université de Paris, Paris, France

3. Correspondence to Varun Warrier (vw260@medschl.cam.ac.uk) or Simon Baron-Cohen (sb205@cam.ac.uk).

4. These authors jointly directed this work

Word count = 3,427

Abstract

Importance

Schizophrenia, educational attainment, and intelligence are all genetically correlated with autism. However, autism is a complex condition, with several different core (such as social communication difficulties and repetitive and restricted behaviour) and associated features (such as IQ and adaptive behaviour) contributing to the underlying heterogeneity. It is unknown to what extent polygenic scores (PGS) for these three phenotypes are associated with the core and associated autism features.

Objective

To investigate the association of PGS for intelligence, educational attainment and schizophrenia on core autism features, IQ and adaptive behaviour in autistic individuals. To further investigate the effects of stratifying by sex and IQ on these associations.

Design

PGS association for the three phenotypes with 12 different autism core and associated features in three cohorts followed by meta-analysis. We additionally conducted sensitivity analyses by stratifying for sex and IQ.

Settings:

Three cross-sectional, multi-centre cohorts comprising autistic with genotype data, and phenotypic information.

Participants

Autistic individuals (total N: 2,512 – 3,681) from three different cohorts: Simons Simplex Collection (N_{max} = 2,233), Autism Genetic Research Exchange (N_{max} = 1,200), and AIMS-2-TRIALS LEAP (N_{max} = 262)

Main outcome measures

Association of PGS for intelligence, educational attainment, and schizophrenia with core autism features, measures of intelligence, and adaptive behaviour in autistic individuals

Results

We identified a similar pattern of correlation among core and associated autism features across all three cohorts. Cluster analyses of these features identified two broad clusters – one consisting of the core features, and another consisting of IQ and adaptive behaviour. PGS for intelligence were only associated with measures of intelligence and adaptive behaviour (e.g. for full-scale IQ, Beta = 0.08, 95%CI = 0.11 – 0.04) for PGS for educational attainment were associated for measures of intelligence, adaptive behaviour, and additionally, non-verbal communication as measured by ADI – a core-autism feature (e.g. for full-scale IQ, Beta = 0.05, 95%CI = 0.08 – 0.02; for ADI non-verbal communication, Beta = 0.05, 95%CI = 0.09 – 0.01). Finally, PGS for schizophrenia were associated with two core autism features: restricted and repetitive behaviour and verbal communication difficulties as measured by the

ADI-R (e.g. for ADI restricted and repetitive behaviour: Beta = 0.06, 95%CI = 0.09 – 0.02). Most of these associations were also significant when restricting it to males only or to individuals with IQ > 70. We find limited evidence for heterogeneity between cohorts.

Conclusion and relevance

We identify specific and different patterns of association between PGS for the three phenotypes and core and associated autism features. This provides greater resolution into the shared genetics between autism and the three phenotypes, and suggests one method to investigate heterogeneity in autism and co-occurring conditions.

Key Questions

Question: Are polygenic scores (PGS) for intelligence, educational attainment, and schizophrenia differentially associated with core autism features, IQ, and adaptive behaviour in autistic individuals?

Findings: Across three cohorts, we identify that PGS for intelligence, educational attainment and schizophrenia are associated with different features in autistic individuals. PGS for intelligence were primarily associated with adaptive behaviour and IQ, PGS for schizophrenia were associated with some core autism features, whilst PGS for educational attainment were associated with IQ, adaptive behaviour, and some core autism features.

Meaning: Our findings suggest that the shared genetics between autism and intelligence, educational attainment, and schizophrenia emerge primarily from specific of core and associated features which are different for the three sets of PGS tested.

Introduction

Autism is a set of complex and heterogeneous neurodevelopmental conditions characterised by difficulties in social communication, unusually repetitive and restricted behaviour and interests, difficulties adjusting to unexpected change, and altered sensory processing (1,2). Autism also frequently co-occurs with other medical, neurodevelopmental, or mental health conditions (3,4). Autism is substantially heritable, with twin and familial heritabilities between 50–90% (5–8). Significant progress has been made in identifying common genetic variants (9) and genes enriched for rare and protein truncating variants associated with autism (10–12). Additionally, genome-wide association studies (GWAS) have identified shared genetics between autism and a number of other phenotypes, including schizophrenia (r_g = 0.24±0.04), educational attainment (r_g = 0.20±0.02), and intelligence (r_g = 0.22±0.03) (9,13).

Autism is unlikely to be a single condition but a composite of multiple, partly dissociable core features (e.g. social interaction and communication difficulties, and restricted interested and repetitive behaviour) and associated features like difficulties in adaptive behaviour and a wide IQ range (14–19). These core and associated features contribute to heterogeneity in autism and are associated with differential outcomes in autistic individuals (20,21). Therefore, investigating the genetics of autism as a single phenotype may provide incomplete information about prognosis and outcomes in autistic individuals.

While schizophrenia, educational attainment, and intelligence are all genetically correlated with autism, we still do not know which core and associated features of autism these phenotypes are genetically correlated with. In other words, are polygenic scores (PGS) for these phenotypes associated with all core and associated autism features, or are they associated with only some of the core and associated features? Finally, how do other sources of heterogeneity such as participants' IQ or sex affect these associations? This can provide better resolution of the shared genetics between various core and associated features and other co-occurring conditions, and identify axes to potentially stratify autistic individuals along to provide better care and support.

To address these questions, we investigated the association of PGS for three well-powered GWAS (intelligence, educational attainment and schizophrenia) with 12 different core and associated autism features in 2,512 – 3,681 autistic individuals from three cohorts. These 12 features represent the core aspects of autism such as restricted and repetitive behaviour and interests, and social interaction and communication difficulties, and associated features such as measures of IQ and adaptive behaviour. We additionally conduct sensitivity analyses by stratifying for males and individuals with IQ > 70, and investigate differences in PGS associations between cohorts.

Methods

Overview of the cohorts

We conducted PGS analyses in three cohorts of autistic individuals: The Simons Simplex Collection (SSC) (22), the Autism Genetic Resource Exchange (AGRE, CHOP sample) (N_{max} = 1,200) (23), and the AIMS-2-TRIALS LEAP sample (N_{max} = 262) (24). Participants in all three cohorts were recruited across multiple sites. For each cohort, we restricted our analyses to autistic individuals who passed the genetic quality control and who had the relevant phenotype, resulting in a total sample size ranging from 2,512 - 3,681. We compare the demographic characteristics of these cohorts in **Table 1**. Further details of these cohorts are provided in the **eMethods 1**.

Insert Table 1 here

Genetic quality control, imputation, and polygenic score generation

Full details of the genetic quality control are provided in **eMethods2.** Briefly, quality control was conducted for each cohort separately, by array. We excluded participants with genotyping rate < 95%, excessive heterozygosity (±3 standard deviations from the mean), who were not of European ancestry as identified using multidimensional scaling, had mismatched genetic and reported sex, and families with Mendelian errors > 10%. SNPs were excluded with genotyping rate < 10%, or deviated from Hardy-Weinberg equilibrium (p-value < $1x10^{-6}$). Imputation was conducted using Michigan Imputation Server (25) using the 1000 genomes Phase 3 v5 as the reference panel (26) (for AGRE and SSC) or using HRC r1.1 2016 reference panel (for AIMS-2-TRIALS). SNPs with minor allele frequency < 1% or an imputation r^2 < 0.6 were excluded.

GWAS used to generate PGS were identified *a priori* using statistical power analyses (27) using https://eagenetics.shinyapps.io/power_website/. Details of parameters used are provided in **eTable 1.** We assumed a SNP heritability of 0.15 for the core and associated autism features and a sample size of 3,000. We conducted power analyses for seven relatively well-powered GWAS that are genetically correlated with autism. The GWAS for autism (9) was excluded as this overlaps with the SSC and AGRE datasets. We subsequently restricted our analyses to only GWAS that had around 80% statistical power at a genetic correlation of 0.25 or greater than 60% power at a genetic correlation of 0.2, which is similar to the genetic correlation between autism and educational attainment, intelligence and schizophrenia. Three phenotypes met our criteria: 1. Educational attainment (28) (N = 766,345, excluding the 23andMe dataset), 2. Intelligence (29) (N = 269,867), and 3. Schizophrenia (30) (56,418 cases and 78,818 controls). All three GWAS are genetically correlated with autism, and PGS for these GWAS are over-transmitted from parents to autistic children (13), suggesting that the observed genetic correlations with autism are not due to other confounding factors such as ascertainment bias.

PGS were generated for three phenotypes using PRSice2 (31), using independent SNPs (clumping r^2 of 0.1 and 250 kb) present in both the GWAS and the testing datasets.

Given the limited testing sample size and the number of tests conducted, we chose *a priori* p-value thresholds that explained maximum variance in the three phenotypes: $p \leq 1$ for the educational attainment and intelligence GWASs (28,29), and $p \leq 0.1$ (30) for the schizophrenia GWAS. The total number of SNPs included are provided in **eTable 2**.

Autism core and associated features

We identified 12 autism core and associated features that are: 1. Widely used in studies related to autism; 2. Are a combination of parent-, self-, other-report and performance-based measures to investigate if reporter status affects the PGS association, 3. Collected in all three cohorts; and 4. Cover a range of core and associated features in autism. Details of each feature is provided in the **eMethods 3**. The core features are:

1. ADOS: Social Affect (ADOS SA)
2. ADOS: Restricted and Repetitive Behaviour domain total score (ADOS RRB)
3. ADI: Communication (verbal) domain total score* (ADI VC)
4. ADI: Communication (nonverbal) domain total score* (ADI NVC)
5. ADI: Restricted and repetitive behaviour domain total score (ADI RRB)
6. ADI: Social domain total score (ADI SOC)
7. Repetitive Behaviour Scale – Revised (RBS)
8. Parent-reported Social Responsiveness Scale: Total raw scores (SRS)

The associated features are :

1. Vineland Adaptive Behaviour Composite Standard scores (VABS)
2. Full-scale IQ (FSIQ)
3. Verbal IQ (VIQ)
4. Non-verbal IQ (NVIQ)

* Verbal and non-verbal communication scores were not separately calculated in the AIMS2-TRIALS-LEAP dataset. We thus used a total communication score and meta-analysed it with the verbal and non-verbal communication scores in the other two cohorts.

Statistical analyses

For each cohort, PGS were regressed against the autism features with sex and the first 15 genetic principal components as covariates in all analyses, with all variables standardised (mean = 0, standard deviation = 1). Additionally, age of participants in the AIMS-2-TRIALS LEAP cohort, and genotype array in the SSC cohort were included as covariates. Age at completion of each measure was only available in the AIMS-2- TRIALS LEAP dataset. IQ type and ADOS module were included as additional covariates. For SSC and AIMS-2-TRIALS LEAP as only unrelated individuals were included, a linear model was used. AGRE includes a large number of multiplex families with multiple autistic individuals per family. Thus, to retain all participants, we used linear mixed effects regression with family ID included as a random intercept to account for relatedness between individuals.

Effect sizes were meta-analysed across the three cohorts using inverse variance weighted meta-analyses with the following formula:

$$w_i = 1/SE_i^2$$
$$SE_{meta} = \sqrt{1/\Sigma_i w_i}$$
$$Beta_{meta} = \Sigma_i \beta_i w_i \, / \Sigma_i w_i$$

Where β_i is the standardized regression coefficient of the PGS, SE_i is the associated standard error, and w_i is the weight. p-values were calculated from Z scores ($Beta_{meta}/SE_{meta}$).

Variance explained was estimated (R^2) by squaring the $Beta_{meta}$. Given the high correlation between the autism features and phenotypes, we used Benjamini-Hochberg False Discovery Rates to correct for multiple testing (q-value < 0.05).

Core and associated autism features are quantitatively and qualitatively different between sexes (32). Further, autistic individuals with co-occurring ID are enriched for rare and *de novo* protein-truncating variants (PTV) (33,34), suggesting that some of the variation in these phenotypes may be attributed to this class of genetic variants in these individuals (35,37). So, we conducted sensitivity analyses by restricting the samples to only males (N = 3,040 – 2,125) or individuals with IQ > 70 (N = 2,319 – 1,801). We did not test for association between FSIQ and VIQ and the three PGS in the AGRE in the two sensitivity analyses, as fewer than 20 individuals had scores on these measures. We followed the same statistical analysis pipeline for the two sensitivity analyses, and identified significant associations using Benjamini-Hochberg FDR correction (q-value < 0.05).

To understand how the features correlated with each other, we conducted Pearson's correlation analyses followed by hierarchical clustering for each cohort (**eMethods 4**).

All analyses were conducted in R. Data and software availability is provided in **eMethods 5.**

Ethics

We received ethical approval to access and analyse de-identified genetic and phenotypic data from the three cohorts from the University of Cambridge Human Biology Research Ethics Committee.

Results

Phenotypic correlation in the three cohorts

Pearson's correlation produced a similar correlation profile across all three cohorts with two clusters – a core autism cluster with all core autism features, and an associated feature cluster with all three IQ measures and VABS (**Figure 1, eFigures 1 – 3, eTables 3 -**

5). The two clusters were largely negatively correlated with each other. Within the autism cluster, hierarchical clustering suggested that instrument type had primacy over domain. For instance, the ADOS subscales clustered closely together, and as did the two parent-report measures (SRS and RBS).

Insert Figure 1 here

Correlation analyses of the correlation coefficients across pairs of features identified high correlations for all three pairs (SSC-AGRE: r = 0.89, p-value < $2x10^{-16}$; SSC-AIMS-2-TRIALS LEAP: r = 0.92, p-value < $2x10^{-16}$; and AGRE – AIMS-2-TRIALS LEAP: r = 0.92, p-value < $2x10^{-16}$) (**Figure 2**).

Insert Figure 2 here

Polygenic score analyses

PGS for intelligence were positively associated with all measures of IQ and VABS with concordant effect direction in all cohorts: FSIQ (Beta = 0.08±0.02, q-value = $6.69x10^{-4}$, R^2 = 0.63%), VIQ (Beta = 0.06±0.02, q-value = $5.21x10^{-3}$, R^2 = 0.37%), NVIQ (Beta = 0.09±0.02, q-value = $3.82x10^{-5}$, R^2 = 0.77%), and VABS (Beta = 0.06±0.02, q-value = $1.91x10^{-2}$) (**Figure 3 and eTable 6**).

Similarly, PGS for educational attainment were positively associated with all measures of IQ and VABS with concordant effect direction in all cohorts: FSIQ (Beta = 0.06±0.02, q-value = $2.36x10^{-3}$, R^2 = 0.30%), VIQ (Beta = 0.06±0.01, q-value = $6.69x10^{-4}$, R^2 = 0.32%), NVIQ (Beta = 0.05±0.01, p-value = $7.57x10^{-3}$, R^2 = 0.21%), and VABS (Beta = 0.04±0.02, q-value = $4.53x10^{-2}$, R^2 = 0.20%) (**Figure 3 and eTable 6)**. There were no statistically significant differences in effect sizes between the PGS for education and intelligence for all measures of IQ.

Interestingly, PGS for educational attainment were also positively associated with scores on the verbal and non-verbal communication subscales of the ADI, though the effect directions were not concordant in the three cohorts (ADI NVC: Beta = 0.05±0.02, q-value = $3.23x10^{-2}$, R^2 = 0.26% ADI VC: Beta = 0.05±0.02, q-value = $2.69x10^{-2}$, R^2 = 0.27%) (**Figure 3 and eTable 6**). Higher scores on the ADI are associated with more difficulties with communication.

PGS for schizophrenia were positively associated with two ADI subscales (ADI RRB: Beta = 0.06±0.02, q-value = $5.21x10^{-3}$, R^2 = 0.39%, ADI NVC: Beta = 0.05±0.02, q-value = $3.23x10^{-2}$, R^2 = 0.32%) (**Figure 3 and eTable 6**). The effect direction was concordant in all three cohorts for ADI RRB but not for ADI NVC (**eTable 6**).

Insert Figure 3 here

Assessing heterogeneity across cohorts and polygenic scores

We next conducted binomial sign test to investigate effect direction concordance between the three cohorts. Formal tests of heterogeneity are likely to be biased given the small number of studies included (36), so we assessed heterogeneity using sign tests assessing effect directions. 32 out of 36 tests had concordant effect direction between the SSC and the AGRE cohorts (p-value = 1.94×10^{-6}, binomial sign test), 21 out of 30 between the SSC and AIMS-2-TRIALS (p-value = 0.04), and 22 out of 30 between the AGRE and the AIMS-2-TRIALS dataset (p-value = 0.01), suggesting largely similar effect directions across cohorts.

We also investigated the concordance of effect direction between the three sets of PGS for the 12 features tested using the meta-analysed effect direction. All 12 tests had concordant effect direction between the PGS for educational attainment and intelligence (p-value = 4.8×10^{-4}, binomial sign test), but only 6 of 12 has concordant effect direction between PGS for schizophrenia and both educational attainment and intelligence (p-value = 1). This is expected given the high genetic correlation between intelligence and educational attainment ($r_g = 0.72 \pm 0.011$), but the low genetic correlation between schizophrenia and intelligence ($r_g = -0.21 \pm 0.025$) and educational attainment ($r_g = 0.01 \pm 0.017$).

Sensitivity analyses

As the AIMS-2-TRIAL LEAP dataset significantly differed from the AGRE and the SSC datasets in terms of fraction of female autistic or autistic individuals with IQ < 70, we conducted two sensitivity analyses (**Methods**). In the males-only analyses, PGS for intelligence were associated with FSIQ and NVIQ (**eFigure 4** and **eTable 7**). PGS for educational attainment were associated with the two ADI communication subscales, RBS, FSIQ, and VIQ. Finally, PGS for SCZ were associated with ADI: RRB.

In the IQ > 70 analyses, PGS for intelligence and educational attainment were associated with FSIQ, VIQ, and NVIQ. Notably, the variance explained in this subset of individuals for these three phenotypes were 2 – 3 times higher than the variance explained in the primary analyse. Additionally, PGS for educational attainment were associated with scores on the VABS, RBS, and ADI: NVC. Finally, PGS for schizophrenia were significantly associated with ADI: RRB and NVIQ (**eFigure 5** and **eTable 8**). In both sensitivity analyses, the effect directions were concordant for all significant associations.

Discussion

Given demographic differences between the three cohorts, we first compared the patterns of phenotypic correlations for the core and associated autism features in the three cohorts, and identified a similar pattern in all three datasets. The core features clustered together, and the measures of intelligence and adaptive behaviour clustered together. This was confirmed by high Pearson's correlation between pairs of features across the three cohorts.

PGS for the three phenotypes were differentially associated with the features tested. Specifically, PGS for educational attainment and intelligence were positively associated with all three measures of IQ and VABS. The shared genetics between VABS and educational attainment and intelligence is unsurprising, given the modest to substantial phenotypic positive correlations between VABS and measures of IQ in the three cohorts (r = 0.36 to 0.68). Sensitivity analyses in autistic individuals with IQ > 70 confirmed the association between the measures of IQ and PGS for educational attainment and intelligence. Previous research has suggested that protein truncating rare and *de novo* variants are negatively associated with measures of IQ and scores on the VABS in autistic individuals (10,33,39), though these sets of variants are enriched in individuals with IQ < 70. In contrast, in our analyses, PGS for educational attainment and intelligence explained a greater proportion of variance in measures of IQ in autistic individuals with IQ > 70 compared to all autistic individuals, suggesting a more prominent role for PGS in measures of IQ in autistic individuals without intellectual disability. Taken together, our analyses suggest that in autistic individuals, similar to the general population (28,29,38), genetic variants across the allelic frequency spectrum contribute to differences in IQ and related measures, but the relative contribution of different sets of genetic variants may differ based on intellectual disability.

PGS for educational attainment were also associated with higher scores on the two ADI communication subdomains (verbal and non-verbal communication), which is in the opposite direction to what's been observed between social and communication difficulties and educational attainment in the general population (r_g = -0.30± 0.11, P = 0.007) (18). Sensitivity analyses restricting to either males or individuals with IQ > 70 also identified significant associations between PGS for educational attainment and ADI communication subdomains with concordant effect direction across all three datasets. This inversion in effect direction between social and communication difficulties and educational attainment in the typical population and autistic individuals must be further investigated. In the two sensitivity analyses PGS for educational attainment were negatively associated with scores on the RBS, though this was only nominally significant in the non-stratified analyses. The variance explained by the PGS for educational attainment for RBS-R in the subgroups was 2 – 2.5x the variance explained in the primary analyses, highlighting, yet again, heterogeneity within the autism spectrum based on IQ and sex.

PGS for schizophrenia, on the other hand, were positively associated with scores on ADI: RRB and ADI: NVC. However, only the association with ADI: RRB was significant in the two sensitivity analyses. In the general population, there is some evidence to suggest that PGS for schizophrenia are associated with social and communication difficulties (40,42). However, to our knowledge, this is the first report linking PGS for schizophrenia with restricted and repetitive behaviour, though RRBs are elevated in individuals with schizophrenia (41).

The sample size for each individual cohort was limited, and the number of cohorts were small excluding the possibility of conducting formal tests of heterogeneity between cohorts. However, analyses of effect direction of the three sets of PGS revealed broad

similarities between cohorts, with notably high concordance in effect direction between the AGRE and the SSC cohorts. This is notable given the substantial differences in demographic characteristics between the cohorts, and potential differences in the SNPs used to generate the polygenic scores for the three cohorts.

Comparing the physical correlations to the effect direction of the PGS further helps to elucidate the differences between the three sets of PGS. Whilst PGS for SCZ is positively associated with the autism core features, and negatively with the associated features, the PGS for educational attainment and IQ were positively associated with most features tested. Notably, PGS for schizophrenia were *significantly associated* with some autism core features, PGS for intelligence were *significantly associated* with all associated features, whereas PGS for educational attainment were *significantly associated* with all associated features and some autism core features suggesting differential association. However, we note that the statistical significance of these associations will be influenced by sample size. In other words, larger sample sizes may identify other associations between the PGS and the features.

Our study also provides a cautionary note about instrument bias which is reflected in the PGS associations. For example, restricted and repetitive behaviours are primarily measured here using RBS, ADI-RRB and ADOS-RRB. The phenotypic correlations between these are low to modest (r : 0.11 – 0.48). Further, SCZ PGS are only associated with ADI-RRB, and PGS for educational attainment are only associated with RBS (in the sensitivity analyses). Together this suggests that idiosyncratic and systematic differences (e.g. reporter bias, number of questions) makes it challenging to compare different core and associated features even if they are meant to capture the same underlying latent trait. Phenotypically, this manifests as modest correlations, and genetically this suggests different underlying genetic architectures for these features, both of which have been demonstrated before (43–45).

Limitations

Whilst the largest study to date, we were statistically well powered to investigate the shared genetics only with three phenotypes, and were not well-powered to explore the co-variance explained at multiple different p-value thresholds in the training datasets. Larger datasets may identify additional significant associations between the current PGS and features tested. Further, whilst well powered, PGS still do not account for all the SNP heritability. We also do not examine the role of rare and low-frequency variants in this study, which are also associated with autism and the three phenotypes (33,38,46). Finally, each many phenotypes included in the study (e.g. SRS and RBS-R) can be further divided into additional subdomains based on factor analyses or evaluating the items included. A comprehensive investigation across all subdomains is warranted, yet is only possible with additional datasets that provide sufficient statistical power.

Conclusions

Core and associated autism features are differentially associated with PGS for intelligence, educational attainment, and schizophrenia. We find limited evidence for heterogeneity between cohorts despite differences in demographic characteristic, though stratifying for males and IQ > 70 identified additional associations with PGS. Our study provides insights into the underlying heterogeneity in autism, and provides greater resolution to the shared genetics between autism and intelligence, educational attainment, and schizophrenia.

Acknowledgements

V.W. is funded by St. Catharine's College, Cambridge. This study was funded by grants to SBC from the Medical Research Council, the Wellcome Trust, the Autism Research Trust, the Templeton World Charity Foundation, and to T.B. from the Institut Pasteur, the CNRS, The Bettencourt-Schueller and the Cognacq-Jay Foundations, the APHP and the Université de Paris. SBC was funded by the Autism Research Trust, the Wellcome Trust, the Templeton World Charitable Foundation, and the NIHR Biomedical Research Centre in Cambridge, during the period of this work. The Medical Research Council (MRC) funded the Cambridge Autism Research Database (CARD) that made this study possible. SBC also received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No 777394. The JU receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA and AUTISM SPEAKS, Autistica, SFARI. His research was also supported by the National Institute of Health Research (NIHR) Applied Research Collaboration East of England (ARC EoE) programme. The views expressed are those of the authors, and not necessarily those of the NIHR, NHS or Department of Health and Social Care. We acknowledge with gratitude the generous support of Drs Dennis and Mireille Gillings in strengthening the collaboration between S.B.-C. and T.B., and between Cambridge University and the Institut Pasteur. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. The AIMS-2-TRIALS LEAP receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA and AUTISM SPEAKS, Autistica, SFARI. A full list of the authors and affiliations in the AIMS-2-TRIALS LEAP group is provided in the Supplementary Information.

References

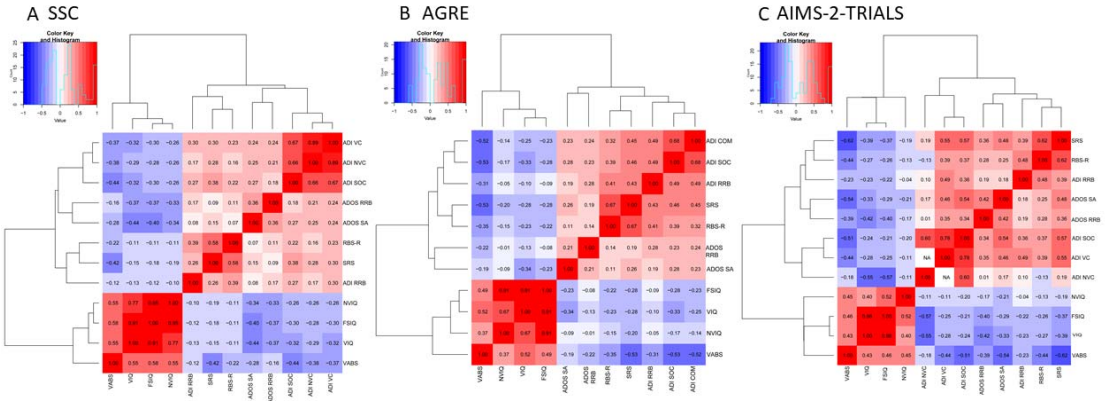
1. Lai M-C, Lombardo M V., Baron-Cohen S (2013): Autism. *Lancet*. https://doi.org/10.1016/S0140-6736(13)61539-1
2. American Psychiatric Association (2013): *Diagnostic and Statistical Manual of Mental Disorders (5th Ed.)*.
3. Lai M-C, Kassee C, Besney R, Bonato S, Hull L, Mandy W, *et al.* (2019): Prevalence of co-occurring mental health diagnoses in the autism population: a systematic review and meta-analysis. *The Lancet Psychiatry* 6: 819–829.
4. Bishop-Fitzpatrick L, Rubenstein E (2019): The physical and mental health of middle aged and older adults on the autism spectrum and the impact of intellectual disability. *Res Autism Spectr Disord* 63: 34–41.

5. Colvert E, Tick B, McEwen F, Stewart C, Curran SR, Woodhouse E, *et al.* (2015): Heritability of autism spectrum disorder in a UK population-based twin sample. *JAMA Psychiatry* 72: 415–23.
6. Tick B, Bolton PF, Happé F, Rutter M, Rijsdijk F (2016): Heritability of autism spectrum disorders: A meta-analysis of twin studies. *J Child Psychol Psychiatry Allied Discip* 57: 585–595.
7. Sandin S, Lichtenstein P, Kuja-Halkola R, Larsson H, Hultman CM, Reichenberg A (2014): The familial risk of autism. *JAMA* 311: 1770–7.
8. Bai D, Yip BHK, Windham GC, Sourander A, Francis R, Yoffe R, *et al.* (2019): Association of Genetic and Environmental Factors With Autism in a 5-Country Cohort. *JAMA Psychiatry* 76: 1035.
9. Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H, *et al.* (2019): Identification of common genetic risk variants for autism spectrum disorder. *Nat Genet* 51: 431–444.
10. Sanders SJ, He X, Willsey AJ, Ercan-Sencicek AG, Samocha KE, Cicek AE, *et al.* (2015): Insights into Autism Spectrum Disorder genomic architecture and biology from 71 risk loci. *Neuron* 87: 1215–33.
11. Satterstrom FK, Kosmicki JA, Wang J, Breen MS, Rubeis S De, An J-Y, *et al.* (2018): Novel genes for autism implicate both excitatory and inhibitory cell lineages in risk. *bioRxiv* 484113.
12. Yuen RKC, Thiruvahindrapuram B, Merico D, Walker S, Tammimies K, Hoang N, *et al.* (2015): Whole-genome sequencing of quartet families with autism spectrum disorder. *Nat Med* 21: 185–91.
13. Weiner DJ, Wigdor EM, Ripke S, Walters RK, Kosmicki JA, Grove J, *et al.* (2017): Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. *Nat Genet*. https://doi.org/10.1038/ng.3863
14. Waterhouse L, London E, Gillberg C (2016): ASD Validity. *Rev J Autism Dev Disord* 3: 302–329.
15. Shuster J, Perry A, Bebko J, Toplak ME (2014): Review of Factor Analytic Studies Examining Symptoms of Autism Spectrum Disorders. *J Autism Dev Disord* 44: 90–110.
16. Mandy WPL, Skuse DH (2008): Research Review: What is the association between the social-communication element of autism and repetitive interests, behaviours and activities? *J Child Psychol Psychiatry* 49: 795–808.
17. Frazier TW, Youngstrom EA, Speer L, Embacher R, Law P, Constantino J, *et al.* (2012): Validation of Proposed DSM-5 Criteria for Autism Spectrum Disorder. *J Am Acad Child Adolesc Psychiatry* 51: 28–40.e3.
18. Warrier V, Toro R, Won H, Leblond CS, Cliquet F, Delorme R, *et al.* (2019): Social and non-social autism symptoms and trait domains are genetically dissociable. *Commun Biol* 2: 328.
19. Greenberg DM, Warrier V, Allison C, Baron-Cohen S (2018): Testing the Empathizing-Systemizing theory of sex differences and the Extreme Male Brain theory of autism in half a million people. *Proc Natl Acad Sci U S A* 201811032.
20. Pickles A, McCauley JB, Pepa LA, Huerta M, Lord C (2020): The adult outcome of children referred for autism: typology and prediction from childhood. *J Child Psychol Psychiatry* jcpp.13180.
21. Anderson DK, Liang JW, Lord C (2014): Predicting young adult outcome among more and less cognitively able individuals with autism spectrum disorders. *J Child Psychol Psychiatry* 55: 485–94.
22. Fischbach GD, Lord C (2010): The Simons Simplex Collection: A Resource for Identification of Autism Genetic Risk Factors. *Neuron* 68: 192–195.

23. Geschwind DH, Sowinski J, Lord C, Iversen P, Shestack J, Jones P, *et al.* (2001): The autism genetic resource exchange: a resource for the study of autism and related neuropsychiatric conditions. *Am J Hum Genet* 69: 463–6.
24. Charman T, Loth E, Tillmann J, Crawley D, Wooldridge C, Goyard D, *et al.* (2017): The EU-AIMS Longitudinal European Autism Project (LEAP): clinical characterisation. *Mol Autism* 8: 27.
25. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, *et al.* (2016): Next-generation genotype imputation service and methods. *Nat Genet* 48: 1284–1287.
26. Gibbs RA, Boerwinkle E, Doddapaneni H, Han Y, Korchina V, Kovar C, *et al.* (2015): A global reference for human genetic variation. *Nature* 526: 68–74.
27. Dudbridge F (2013): Power and Predictive Accuracy of Polygenic Risk Scores ((N. R. Wray, editor)). *PLoS Genet* 9: e1003348.
28. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, *et al.* (2018): Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet* 50: 1112–1121.
29. Savage JE, Jansen PR, Stringer S, Watanabe K, Bryois J, de Leeuw CA, *et al.* (2018): Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat Genet* 1.
30. Lam M, Chen C-Y, Li Z, Martin AR, Bryois J, Ma X, *et al.* (2019): Comparative genetic architectures of schizophrenia in East Asian and European populations. *Nat Genet* 1–9.
31. Euesden J, Lewis CM, O'Reilly PF (2015): PRSice: Polygenic Risk Score software. *Bioinformatics* 31: 1466–1468.
32. Lai M-C, Lombardo M V., Auyeung B, Chakrabarti B, Baron-Cohen S (2015): Sex/gender differences and autism: setting the scene for future research. *J Am Acad Child Adolesc Psychiatry* 54: 11–24.
33. Satterstrom F, Kosmicki J, J W, MS B, S DR, JY A, *et al.* (2020): Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell* 180. https://doi.org/10.1016/J.CELL.2019.12.036
34. Sanders SJ, He X, Willsey AJ, Devlin B, Roeder K, State MW, *et al.* (2015): Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci Article Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron* 87: 1215–1233.
35. Bishop SL, Farmer C, Bal V, Robinson EB, Willsey AJ, Werling DM, *et al.* (2017): Identification of Developmental and Behavioral Markers Associated With Genetic Abnormalities in Autism Spectrum Disorder. *Am J Psychiatry* 174: 576–585.
36. von Hippel PT (2015): The heterogeneity statistic I(2) can be biased in small meta-analyses. *BMC Med Res Methodol* 15: 35.
37. A B, N V, AM K, C, Lord, AE L, M W, I I (2018): Damaging de novo mutations diminish motor skills in children on the autism spectrum. *Proc Natl Acad Sci U S A* 115. https://doi.org/10.1073/PNAS.1715427115
38. Ganna A, Genovese G, Howrigan DP, Byrnes A, Kurki MI, Zekavat SM, *et al.* (2016): Ultra-rare disruptive and damaging mutations influence educational attainment in the general population. *Nat Neurosci* 19: 1563–1565.
39. Robinson EB, St Pourcain B, Anttila V, Kosmicki JA, Bulik-Sullivan BK, Grove J, *et al.* (2016): Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. *Nat Genet* 48: 552–5.
40. St Pourcain B, Robinson EB, Anttila V, Sullivan BB, Maller J, Golding J, *et al.* (2017): ASD and schizophrenia show distinct developmental profiles in common genetic overlap with population-based social communication difficulties. *Mol Psychiatry*. https://doi.org/10.1038/mp.2016.198

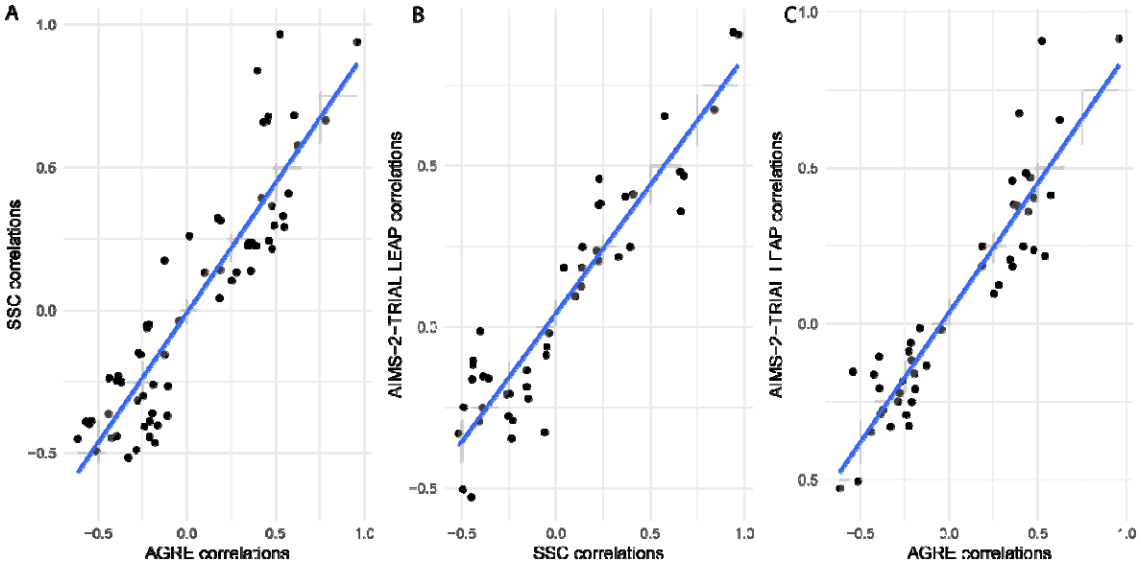
487 41. Evans DW, Uljarević M, Lusk LG, Loth E, Frazier T (2017): Development of Two
488 Dimensional Measures of Restricted and Repetitive Behavior in Parents and Children. *J*
489 *Am Acad Child Adolesc Psychiatry* 56: 51–58.
490 42. Riglin L, Collishaw S, Richards A, Thapar AK, Maughan B, O'Donovan MC, Thapar A
491 (2017): Schizophrenia risk alleles and neurodevelopmental outcomes in childhood: a
492 population-based cohort study. *The lancet Psychiatry* 4: 57–62.
493 43. Mallard TT, Linnér RK, Okbay A, Grotzinger AD, Vlaming R de, Meddens SFW, *et al.*
494 (2019): Not just one p: Multivariate GWAS of psychiatric disorders and their cardinal
495 symptoms reveal two dimensions of cross-cutting genetic liabilities. *bioRxiv* 603134.
496 44. Knott F, Dunlop A-W, Mackay T (2006): Living with ASD: how do children and their
497 parents assess their difficulties with social interaction and understanding? *Autism* 10:
498 609–17.
499 45. Johnson SA, Filliter JH, Murphy RR (2009): Discrepancies between self- and parent-
500 perceptions of autistic traits and empathy in high functioning children and adolescents
501 on the autism spectrum. *J Autism Dev Disord* 39: 1706–14.
502 46. Singh T, Kurki MI, Curtis D, Purcell SM, Crooks L, McRae J, *et al.* (2016): Rare loss-of-
503 function variants in SETD1A are associated with schizophrenia and developmental
504 disorders. *Nat Neurosci* 19: 571–577.
505

Figure 1: Heatmap of the phenotypic correlations in the three cohorts



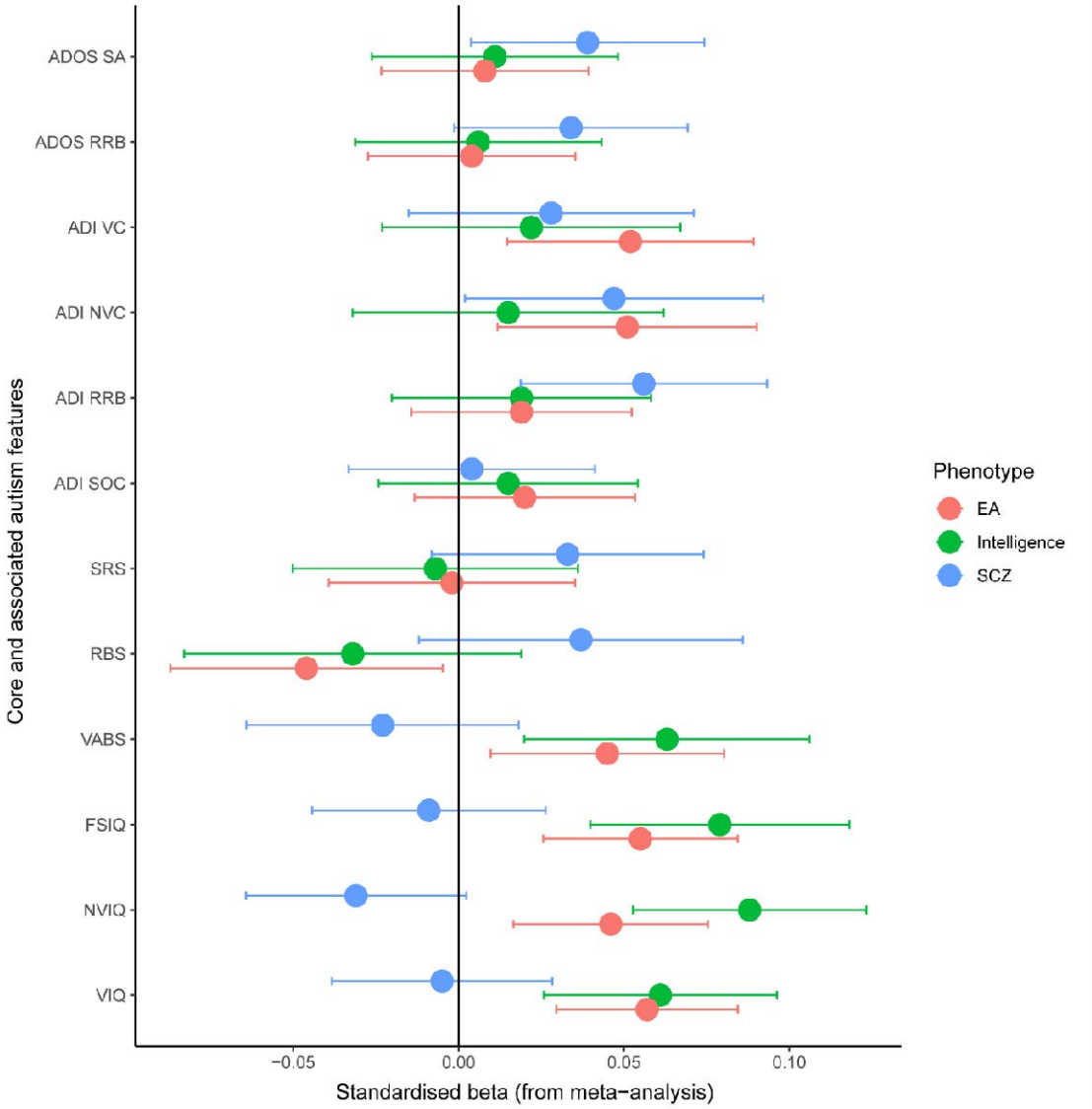
This figure provides the correlation heatmap for the correlations for autism core features, IQ, and adaptive behaviour in the: A. SSC; B: AGRE; and C: AIMS-2-TRIALS LEAP cohorts. Hierarchical clustering tree is also provided for each of the plots. Red colour indicates a positive correlation, while blue indicates a negative correlation. Individual plots for each cohort with greater resolution is provided in eFigures 1 – 3. Separate verbal communication and non-verbal communication ADI scores (ADI VC and ADI NVC) were not available for the AIMS-2-TRIALS LEAP, so we used a composite ADI communication score (ADI COM). Additionally, we did not identify individuals who had both ADI VC and ADI NVC scores in the AGRE cohort, and hence could not calculate the correlations between these two subscales in the AGRE cohort. The features are: ADOS - Social Affect (ADOS SA); ADOS - Restricted and Repetitive Behaviour domain total score (ADOS RRB); ADI - Communication (verbal) domain total score (ADI VC); ADI - Communication (nonverbal) domain total score (ADI NVC); ADI - Restricted and repetitive behaviour domain total score (ADI RRB); ADI - Social domain total score (ADI SOC); Repetitive Behaviour Scale – Revised (RBS-R); Vineland Adaptive Behaviour Composite Standard scores (VABS); Parent-reported Social Responsiveness Scale - Total raw scores (SRS); Full-scale IQ (FSIQ); Verbal IQ (VIQ); and Non-verbal IQ (NVIQ).

Figure 2: Correlations of correlations among pairs of autism core and associated features across the three cohorts



This figure provides the correlation between the correlation coefficients of pairs of core and associated autism features in two datasets. A: Between the SSC and AGRE (r = 0.89, p-value <2x10^{-16}); B: Between AIMS-2-TRIALS LEAP and SSC (r = 0.92, p-value < 2x10^{-16}); and C: Between AIMS-2-TRIALS and AGRE (r = 0.92, p-value < 2x10^{-16}). Pearson's correlation coefficient was calculated. The shaded portion of the line indicates the 95% confidence interval for the correlation.

Figure 3: Polygenic score association of the three phenotypes with the autism core features, measures of IQ, and adaptive behaviour



This figure provides the meta-analysed regression Beta and the associated 95% confidence intervals for the three phenotypes (EA = educational attainment in red, SCZ = schizophrenia in blue, and intelligence in green) against the core autism features, measures of IQ, and adaptive behaviour. These are: ADOS - Social Affect (ADOS SA); ADOS - Restricted and Repetitive Behaviour domain total score (ADOS RRB); ADI - Communication (verbal) domain total score (ADI VC); ADI - Communication (nonverbal) domain total score (ADI NVC); ADI - Restricted and repetitive behaviour domain total score (ADI RRB); ADI - Social domain total score (ADI SOC); Repetitive Behaviour Scale – Revised (RBS); Vineland Adaptive Behaviour Composite Standard scores (VABS); Parent-reported Social Responsiveness Scale - Total raw scores (SRS); Full-scale IQ (FSIQ); Verbal IQ (VIQ); and Non-verbal IQ (NVIQ).