Social and non-social autism symptoms and trait domains are genetically dissociable

Varun Warrier¹, Roberto Toro², Hyejung Won³, Claire S. Leblond², Freddy Cliquet², Richard Delorme²,4, Ward De Witte⁵, Janita Bralten⁵,6, Bhismadev Chakrabarti¹,7, Anders D. Børglum⁸,9,10, Jakob Grove⁸,9,10,11, Geert Poelmans⁵, the 23andMe Research Team#, David A. Hinds¹²,13, Thomas Bourgeron²,13 & Simon Baron-Cohen¹,13

The core diagnostic criteria for autism comprise two symptom domains – social and communication difficulties, and unusually repetitive and restricted behaviour, interests and activities. There is some evidence to suggest that these two domains are dissociable, though this hypothesis has not yet been tested using molecular genetics. We test this using a genome-wide association study (N = 51,564) of a non-social trait related to autism, systemising, defined as the drive to analyse and build systems. We demonstrate that systemising is heritable and genetically correlated with autism. In contrast, we do not identify significant genetic correlations between social autistic traits and systemising. Supporting this, polygenic scores for systemising are significantly and positively associated with restricted and repetitive behaviour but not with social difficulties in autistic individuals. These findings strongly suggest that the two core domains of autism are genetically dissociable, and point at how to fractionate the genetics of autism.

¹Autism Research Centre, Department of Psychiatry, University of Cambridge, Cambridgeshire, UK. ²Human Genetics and Cognitive Functions, Institut Pasteur, UMR3571 CNRS, Université de Paris, Paris, France. ³Department of Genetics and Neuroscience Center, University of North Carolina, Chapel Hill, NC 27599, USA. ⁴Child and Adolescent Psychiatry Department, Robert Debré Hospital, Paris, France. ⁵Department of Human Genetics, Radboud University Medical Center, Nijmegen, The Netherlands. ⁶Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands. ⁷Centre for Autism, School of Psychology and Clinical Language Sciences, University of Reading, Reading, UK. ⁸The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Aarhus, Denmark. ⁹Centre for Integrative Sequencing, iSEQ, Aarhus University, Aarhus, Denmark. ¹⁰Department of Biomedicine – Human Genetics, Aarhus University, Aarhus, Denmark. ¹¹Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark. ¹²23andMe Inc., Mountain View, CA 94043, USA. ¹³These authors contributed equally: David A. Hinds, Thomas Bourgeron, Simon Baron-Cohen. #Members of the 23andMe Research Team are listed at the end of the paper. Correspondence and requests for materials should be addressed to V.W. (email: vw260@medschl.cam.ac.uk) or to S.B.-C. (email: sb205@cam.ac.uk)
The core diagnostic criteria of autism comprises two symptom domains: difficulties in social interactions and communication (the social domain), and unusually repetitive and restricted behaviour and stereotyped interests (the non-social domain)\(^1\). Multiple lines of evidence suggest that these two domains are dissociable\(^2\). First, factor and principal component analysis of autism and autistic traits have predominantly identified two factors—a social and a non-social factor\(^4\)–\(^9\). Second, investigations of autistic traits in large cohorts have demonstrated a positive phenotypic correlation between different social traits and different non-social traits separately, but only a limited correlation between social and non-social traits\(^8\)–\(^12\). Third, twin genetic correlations between social and non-social symptom domains in autism are low, although both social and non-social trait domains are highly heritable in neurotypical\(^13\)–\(^14\) or autistic twins\(^15\). Fourth, difficulties in social and non-social domains can occur independently of each other\(^6\),\(^17\), which has been used to subgroup individuals on the spectrum based on the two domains\(^18\). This suggests that the genetic and phenotypic architecture of autism consists of at least two broadly dissociable domains. This has implications for genetic, biological, and clinical studies of autism, since most studies have investigated autism as if it is a unitary condition\(^3\). The idea that social and non-social symptom domains are dissociable is unsurprising given their very different nature, and very different underlying neurology and cognitive processes: one related to interpreting animate motion and mental states (theory of mind) and the other related to recognising inanimate objects, events or patterns (systemising)\(^3\). Nevertheless, a diagnosis of autism is only given when the social and non-social symptom domains cluster together.

However, to date, there has been limited molecular genetic evidence in support of this dissociability hypothesis, partly due to the limited large-scale research on the genetics of social and non-social domains. Most genetic research into the social and non-social domains has been primarily through linkage and genome-wide association studies (GWAS) in relatively small samples of autistic individuals and the general population \((N < 5K)\)\(^19\)–\(^25\). This has precluded a detailed molecular genetic investigation of the social and non-social domains associated with autism. Given currently available sample sizes with phenotypic information, investigating the genetics of the social and non-social domains in autistic individuals is difficult. However, several studies have demonstrated that the underlying liability for autism is normally distributed in the general population\(^26\)–\(^29\). Factor analyses have failed to identify discontinuities between clinical autism and autistic traits in the general population\(^30\). Autistic traits are heritable\(^31\)–\(^33\), are elevated in family members of autistic individuals compared to the general population\(^34\),\(^35\), and are transmitted inter-generationally\(^36\),\(^37\). Factor analysis of autistic traits measures have also identified two different factors in both the general population and autistic individuals—one linked to the social domain, and another linked to the non-social domain, mirroring the factor structure of clinical autism domains\(^6\),\(^9\),\(^30\),\(^38\). Studies have further demonstrated moderate to high shared genetics between the extremes of the liability distribution and the rest of the distribution\(^14\),\(^39\)–\(^41\). One twin study investigated the bivariate genetic correlation between research and clinical autism diagnosis and autistic traits and identified high genetic correlations \((0.7 < r_g < 0.89)\)\(^42\). Validating this, studies have identified modest shared genetics between autism and autistic traits\(^43\)–\(^45\). Taken together, there is considerable evidence to suggest that autism represents the extreme end of the autistic traits continuum.

While a few studies have investigated the genetics of traits contributing to the social domains such as social and communication difficulties\(^19\),\(^44\)–\(^45\), empathy\(^46\), and emotion recognition\(^47\), there have been limited studies investigating the genetics of the non-social domain\(^25\),\(^48\). Neither of these studies have replicably identified significant variants associated with the non-social domain, primarily because of the relatively modest sample sizes of the GWAS. An alternate approach is to investigate the genetics of non-social traits related to autism in the typical population, maximising the sample size. To better understand the genetics of a non-social trait related to autism, we investigate the genetics of systemising measured using a 75-item well validated, self-report measure called the Systemising Quotient-Revised (SQ-R) (see ‘Methods’ section). Systemising involves identifying input–operation–output (or if-and-then) relationships in order to analyse and build systems, and to understand the laws that govern specific systems\(^49\). The hyper-systemising theory of autism proposes that autistic individuals, on average, have superior attention to detail, and a stronger drive to systemise compared to individuals in the general population\(^50\). This has been validated in several studies\(^50\),\(^51\) including a recent study in more 650,000 individuals, including 36,000 autistic individuals\(^12\). Several lines of evidence suggest that autistic individuals have at least intact if not superior systemising. The idea was noted in the earliest papers describing autism by both Asperger\(^52\) and Kanner\(^53\), although these early papers do not use the term ‘systemising’ but instead comment on strong interests in pattern recognition, the need for order and predictability, excellent memory for facts, and a strong focus on objects and understanding how things work. Further, autistic adults, on average, score higher on the SQ-R compared to individuals in the general population\(^10\),\(^51\), a profile also observed in autistic children\(^54\). Several items in the SQ-R specifically measure circumscribed interests and insistence on sameness, two of the items mentioned in the DSM-5, and several of these items map onto items on the Autism Spectrum Quotient (AQ), a well validated measure of autistic traits\(^27\) (see Supplementary Note). Because systems follow rules, they repeat. A fascination with systems may thus manifest as unusually repetitive behaviour. And because systems depend on precise variables, a fascination with systems may also manifest as unusually narrow interests in autism.

The present study has two aims: first, to investigate the genetic architecture of a non-social trait linked to autism (systemising); and second, to investigate whether social and non-social traits related to autism, measured in the general population, are genetically dissociable.

**Results**

**GWAS results.** We first conducted a GWAS of systemising \((N = 51,564)\) measured using the SQ-R. Following this, and using data from GWAS of social traits genetically correlated with autism (GWAS of self-reported empathy \((N = 46,861)\)\(^46\), and GWAS of social relationship satisfaction\(^35\) measured using friendship \((N_{\text{effective}} = 164,112)\) and family relationship \((N_{\text{effective}} = 158,116)\) satisfaction scales), we investigated whether the social and non-social traits related to autism are genetically dissociable in the general population. A flow chart of the study design is provided in Fig. 1.

Systemising was measured in the 23andMe sample \((N = 51,564)\) using scores from the SQ-R\(^10\). Scores on the SQ-R were normally distributed, with a mean of 71 ± 21 out of 150. As hypothesised based on previous research\(^10\),\(^12\),\(^23\), males \((76.5 ± 20)\), on average, scored higher than females \((65.4 ± 20.6)\) \((P < 0.001, \text{Cohen’s } d = 0.54\), Supplementary Fig. 1). Given the significant sex differences in scores, we conducted a non-stratified and sex-stratified GWAS for the SQ-R. Genome-wide association analyses identified three significant loci (Fig. 2, Supplementary Data 1 and
Supplementary Fig. 2). Two of these were significant in the non-stratified GWAS: rs4146336 on chromosome 3 \((P = 2.58 \times 10^{-8})\) and rs1559586 on chromosome 18 \((P = 4.78 \times 10^{-8})\). The third significant locus was in the males-only GWAS \((rs8005092 on chromosome 14, P = 3.74 \times 10^{-8})\). rs8005092 and rs1559586 lie in regions of high genetic recombination. Linkage-disequilibrium score regression (LDSR) intercept suggested that there was minimal inflation due to population stratification (Fig. 2). Fine-mapping of the three regions identified 14 credible SNPs (see 'Methods' section). None of the SNPs overlapped with fetal brain eQTL. However, two of these SNPs mapped onto two genes—LSAMP and PTMAP8, both of chromosome 3—using chromatin interaction data in the fetal brain. Of these, LSAMP is a neuronal adhesion molecule in the limbic system of the developing brain. In addition, gene-based analysis identified four genes and rs1559586 on chromosome 18 \((ZSWIM6, LSAMP, FUT8, ZNF574)\), Of these, mutations in ZSWIM6 cause a neurodevelopmental disorder with, in some cases, co-morbid autism and unusually repetitive movements and behaviour \(^{56}\). As supporting analyses, we investigated the direction of effect for all independent SNPs with \(P < 1 \times 10^{-6}\) in the non-stratified SQ-R GWAS in GWAS of autism \(^{57}\), educational attainment and cognitive aptitude \(^{58}\), and cognitive aptitude \(^{59}\). Five out of six SNPs tested had concordant effect direction in the GWAS for educational attainment and GWAS for cognitive aptitude \((P = 0.21\), two-sided binomial sign test for each comparison). Similarly, four out of five SNPs tested had concordant effect direction in the GWAS for autism (Supplementary Data 3a) \((P = 0.37\), two-sided binomial sign test). For these three phenotypes, we additionally assessed effect direction concordance using binomial sign test at least stringent \(P\)-value thresholds in the SQ-R GWAS, after LD-based clumping \((P < 1, 0.5, 0.1\) and \(1 \times 10^{-4}\)\). Binomial sign test was statistically significant at three of the four \(P\)-value thresholds \((P = 1, 0.5\) and 0.1) for all three phenotypes but not statistically significant at \(P = 1\times 10^{-4}\), presumably due to the low statistical power (Supplementary Data 3b). In addition, we tested effect direction concordance \((P < 1 \times 10^{-6}\) in a GWAS \((N = 1981)\) of 'insistence on sameness', a phenotype similar to systemising (see 'Methods' section). Four out of five SNPs had a concordant effect direction including the two SNPs with \(P < 5 \times 10^{-8}\) in the non-stratified SQ-R GWAS \((P = 0.37\), two-sided binomial sign test).

**Genetic correlation between the SQ-R and other phenotypes.** Additive SNP-based heritability \((\hat{h}_{SNP}^2)\) calculated using LDSR was 0.12 ± 0.012 for the SQ-R \((P = 1.2 \times 10^{-20})\). Despite small but significant sex differences in the SQ-R scores, there was no significant difference in \(h_{SNP}^2\) between males and females \((P = 0.34)\) (Supplementary Fig. 3 and Supplementary Data 4), which was strengthened by the high genetic correlation between males and females \((1 ± 0.17; P = 3.91 \times 10^{-10})\), suggesting a similar polygenic architecture between sexes. The per-SNP effect for the most significant SNPs was small, suggesting a highly polygenic architecture \((R^2 = 0.001–0.0002\%\), after correcting for winner’s curse, Supplementary Data 5). Partitioned heritability for functional categories identified significant enrichment for evolutionary conserved regions, transcription start sites, fetal DNase hyper-sensitivity sites, and H3 lysine 27 acetylation (H3K27ac), suggesting a prominent role for regulatory and conserved genomic regions in systemising (Supplementary Data 6). Partitioning heritability based on tissue-specific active chromatin marks identified a significant enrichment for brain specific chromatin signatures. Notably, this enrichment was significant in both adult and fetal brain specific active chromatin marks (Supplementary Data 7 and Supplementary Fig. 4). Enrichment for genes expressed in the brain was high.

**Fig. 1** Schematic diagram of the study. We conducted a GWAS of the SQ-R \((N = 51,564)\) and quantified SNP heritability \((\hat{h}_{SNP}^2)\), quantified genetic correlations with multiple phenotypes, and conducted polygenic score analyses. In addition, we conducted sex-stratified GWAS of the SQ-R, and investigated \(h_{SNP}^2\) within sex and genetic correlation between males and females. Finally, we investigated the clustering of all phenotypes that are genetically correlated with autism, and whether the social and the non-social phenotypes associated with autism are genetically correlated with autism.

| Phenotype | Census | Sex-stratified | Polygenic score analyses | Clustering of all phenotypes | Genetic correlation |
|-----------|--------|----------------|--------------------------|-----------------------------|-------------------|
| SQ-R GWAS | 51,564 | 26,063 | 25,501 | N/A | N/A |
| Autism GWAS | N/A | N/A | N/A | N/A | N/A |

| Phenotype | Census | H² SNP | Genetic correlation | Polygenic score analyses | Clustering of all phenotypes | Genetic correlation |
|-----------|--------|--------|-------------------|--------------------------|-----------------------------|-------------------|
| SQ-R GWAS | 51,564 | N/A | N/A | N/A | N/A | N/A |
| Autism GWAS | N/A | N/A | N/A | N/A | N/A | N/A |
We validated this using genomic structural equation modelling (GSEM) (see ‘Methods’ section) using both educational attainment and cognitive aptitude (Fig. 3c). Further, the SQ-R was not genetically correlated with any of the social measures related to autism—friendship and family relationship satisfaction, scores on a self-report measure of empathy (the EQ), and the scores on the Social and Communication Disorders Checklist (SCDC), which is a measure of social and communication difficulties (see Supplementary Note for how these traits map onto social domains in autism). Estimates of genetic correlations between SQ-R scores and the various social traits are also small, suggesting that there is limited shared genetics between social autism traits and the SQ-R.

Genetic correlations between social/non-social traits and psychiatric conditions. To understand the genetic relationship between the SQ-R and autism in a broader context, we evaluated the genetic correlations between multiple phenotypes with evidence of significant genetic correlation with autism (15 phenotypes in total, see ‘Methods’ section for a list of phenotypes included). Clustering highlighted three broad clusters: a social cluster, a psychiatric cluster, and an intelligence cluster (Fig. 4a and Supplementary Tables 11 and 12). The SQ-R clusters closely with measures of intelligence, but while educational attainment...
Fig. 3 Genetic correlation between the SQ-R and other phenotypes, and GWIS and GSEM estimates between SQ, educational attainment and cognitive aptitude. 

a Genetic correlations between the SQ-R and multiple other phenotypes provided. The bars represent 95% confidence intervals. Sample sizes and PMID are provided in Supplementary Data 9. The following genetic correlations were significant after Bonferroni correction: autism ($r_g = 0.26 \pm 0.06; P = 3.35 \times 10^{-5}, N = 46,350$), years of schooling ($r_g = 0.13 \pm 0.03; P = 4.73 \times 10^{-5}, N = 293,723$), college completion ($r_g = 0.18 \pm 0.05; P = 1.30 \times 10^{-3}, N = 95,427$), and cognitive aptitude ($r_g = 0.19 \pm 0.04; P = 2.35 \times 10^{-5}, N = 78,308$).

b Results of the GWIS analysis. Red lines represent genetic correlation with the SQ-R, blue lines represent genetic correlations with the SQ-R independent of the genetic effects of educational attainment. The bars represent 95% confidence intervals.

c Path diagrams providing the results of the standardised SEM models to investigate whether the SQ-R is genetically correlated with autism independent of the genetic effects of cognitive aptitude ($CA_g$) and educational attainment ($EA_g$). GWIS genome-wide inferred statistics, GSEM genomic structural equation modelling.
**Fig. 4** Genetic correlogram of autism and related traits, and genetic correlations between social and non-social traits and multiple psychiatric conditions. 

**a** Correlogram of genetic correlations between all phenotypes that are genetically correlated with autism. Please note the upper and lower triangle are identical. Asterisk (provided only in the lower triangle) represents significant correlations after Bonferroni correction. Genetic correlations have been clustered using hierarchical clustering. Colour provides the magnitude of genetic correlation.

**b** Genetic correlation between empathy, friendship satisfaction, and systemising with nine psychiatric conditions. Only autism was significantly genetically correlated with all three phenotypes. Full results are present in Supplementary Data 11.
and cognitive aptitude are significantly genetically correlated with multiple social traits and psychiatric conditions, the SQ-R is only genetically correlated with autism.

Given that the two major domains of autism as identified by the DSM-5 are persistent difficulties in social interaction and communication and unusually restrictive, stereotyped, and repetitive interests\(^6\), we hypothesised that the combination of significant negative genetic correlation with social traits (friendship satisfaction and empathy) and significant positive genetic correlation with SQ-R would be uniquely associated with autism (see ‘Methods’ section). Indeed, across the nine psychiatric conditions for which we had summary GWAS statistics, this combination was uniquely observed for autism (Fig. 4b, Supplementary Data 13).

**Validation in additional cohorts.** Given that our current analysis focussed on the general population, we sought to investigate whether polygenic scores from the SQ-R were associated with social and non-social autism traits in 2221 autistic individuals from the Simons simplex collection (see ‘Methods’ section). We hypothesised that SQ-R may be significantly associated with the non-social domain in autism, but not associated with the social domain in autism. Polygenic scores for SQ-R were significantly associated with scores on the Repetitive Behaviour Scale-Revised (RBS-R) (\(\beta = 0.052 \pm 0.02, P = 0.013\)), but not on the social and communication subscale of ADOS-G (\(\beta = -0.00099 \pm 0.018, P = 0.95\)) after adjusting for multiple test (Bonferroni alpha = 0.025). We validated this in 426 additional individuals of which 401 had a diagnosis of autism with RBS-R scores from the EU-AIMS LEAP, AGRE, and Paris cohorts. Here, we identified a concordant effect direction for polygenic scores of the SQ-R (\(\beta = 0.02 \pm 0.05, P = 0.65\)), although the results were not significant potentially due to the small sample size. Inverse-variance meta-analysis of the discovery and the validation cohorts marginally improved the significance of the association (\(\beta = 0.047 \pm 0.018, P = 0.010\)), and the results remained statistically significant (Bonferroni alpha = 0.025). In a separate sample of 475 autistic individuals from the AGRE cohort, polygenic scores for the SQ-R were not associated with the social and communication subscale of ADOS-G (\(\beta = -0.046 \pm 0.04, P = 0.24\)). Meta-analysis of the two cohorts did not produce a statistically significant result (\(\beta = -0.008 \pm 0.016, P = 0.60\)) (see Power calculations in the Supplementary Note). We note that the lack of association between the polygenic scores for the SQ-R and the ADOS-G social and communication subscale is not indicative of absence of shared genetics, but rather indicative of lower shared genetics between the SQ-R and the ADOS-G social and communication subscale than that between the RBS-R and the SQ-R.

Finally, to further validate the results in autistic individuals, we conducted bivariate genetic correlations on scores on the RBS-R and the ADOS-G social and communication subscale in 2989 individuals from the SSC, AGRE, EU-AIMS LEAP and Paris cohorts (2964 autistic individuals). Both the RBS-R (\(h^2_{\text{SNP}} = 0.11 \pm 0.11, P = 0.15\)) and the ADOS-G social and communication subscale (\(h^2_{\text{SNP}} = 0.26 \pm 0.10, P = 0.004\)) had modest \(h^2_{\text{SNP}}\), though only the latter was statistically significant. We identified a small genetic correlation (\(r_g = 0.15 \pm 0.46, P = 0.74\)), which was not statistically different from 0. Given the small sample size, the genetic correlation is unlikely to be statistically significant. However, the effect was small and statistically less than 1 (\(P = 0.034\), one-tailed t-test).

**Discussion**

Here we present, to our knowledge, the largest GWAS of a non-social trait related to autism in the general population—systemising, measured using the SQ-R. We demonstrate that systemising is heritable and genetically correlated with autism. Associated loci are enriched in genomic regions containing brain chromatin signatures and we identify three genome-wide significant loci, but these must be replicated in an independent cohort. Despite the modest sample size, our GWAS is well-powered to investigate genetic correlations between various phenotypes including social traits related to autism, as the \(Z\)-score of the \(h^2_{\text{SNP}}\) is above the recommended threshold of four\(^6\). We identify high sign concordance of the top SNPs in genetically correlated traits, enrichment for active chromatin marks in fetal and adult brain, and significant polygenic score association with the RBS-R. Polygenic score analysis suggests that the shared genetics between systemising and the non-social domain of autism is considerably higher than the shared genetics between systemising and the social domain of autism. In addition, using a smaller sample of autistic individuals, we provide preliminary evidence that the social and non-social domains in autistic individuals have low shared genetics. Our results highlight the need to collect deeper clinical and cognitive information in autistic individuals to better understand the phenotypic heterogeneity in autism.

Most studies model autism and autistic traits as a single phenotype. This has likely arisen because of the difficulties in recruiting and phenotyping sufficient numbers of autistic people. Our study suggests that both in the general population and in autistic individuals, social and non-social autistic traits and symptom domains are genetically dissociable. This may to some extent explain why, compared to GWAS of other psychiatric conditions of roughly similar sample sizes\(^57,63-65\), the largest GWAS of autism to date has identified fewer loci. One possible explanation is statistical signal-attenuation because of the underlying heterogeneity. However, this does not necessarily suggest that systemising, or the other individual trait domains are less complex. For instance, we observe similar \(h^2_{\text{SNP}}\) for SQ-R, self-reported empathy\(^46\), and the largest and most recent GWAS of autism\(^57\).

It is important to investigate whether these domains are dissociable in a larger cohort of autistic individuals and identify potential convergence of the two domains in gene expression networks in the developing brain. Our results confirm the need to rethink our understanding of autism as existing along a single dimension\(^3,66\). We hypothesise that the dissociation of the two domains will extend to other research modalities in studies of autism and autistic traits. It is important to note that, while our results demonstrate two broadly dissociable autistic trait domains in the general population and in autistic individuals, more research is needed to identify other potentially dissociable domains and to investigate whether this dissociability is driven by different designs of phenotypic instruments (e.g. self-report vs informant report). For example, our research does not make a distinction between communication and social interaction abilities, or between sensory difficulties and repetitive behaviours, and future molecular genetic studies may identify varying levels of overlap between these domains. The same principle applies to other research modalities (neuroimaging, cognitive studies, hormonal assays, etc.) investigating the biology of autism and autistic traits. These different symptom domains of autism may contribute to different co-morbidities. Our results identify shared genetics between the social traits related to autism and psychiatric conditions such as schizophrenia and depression, but limited shared genetics between the SQ-R and these conditions.

**Methods**

**Participants.** The current study included participants from 23andMe (primary GWAS - SQ-R), from ALSPAC (GWAS of scores on the Social and Communication Disorders Checklist (SCDC)) and autistic individuals from the Simons
**Simplex Collection (SSC), the Autism Genetic Resource Exchange ( AGRE), and the EU-AIMS LEAP and PARIS cohorts.**

**23andMe participants.** Research participants in the GWAS of the SQ-R were from 23andMe and are described in detail elsewhere67,68. All participants provided informed consent and answered surveys online according to a human subjects’ research protocol, which was reviewed and approved by Ethical & Independent Review Services, an external AAHRPP-accredited private institutional review board (http://www.eandireview.com). All participants completed the online version of the SQ-R during the 23andMe parental portal. Only participants who were primarily of European ancestry (97% European Ancestry) were selected for the analysis using existing methods89. Unrelated individuals were selected using a segmental identity-by-descent algorithm90. A total of 51,564 participants completed the SQ-R (males=26,063, and females=25,501).

**ALSPAC participants.** ALSPAC is a longitudinal cohort which recruited pregnant mothers in the Avon region of the UK. The ALSPAC cohort comprises 14,541 initial pregnancies from women in Avon resulting in a total of 13,988 children who were alive at 1 year of age. Children were enrolled in additional phases, described in greater detail elsewhere4. This study received ethical approval from the ALSPAC Law-and-Ethics Committee, and the Cambridge Human Biology Research Ethics Committee. Written informed consent was obtained from parent or a responsible legal guardian for the child to participate. Assent was obtained from the child participants where possible. We conducted a GWAS of scores on the SCDC in 5,421 individuals from ALSPAC.

**Other cohorts.** We included data from four cohorts to conduct polygenic score and bivariate genetic correlation analyses. The SSC (n=2,221 unrelated autistic individuals) consists of simplex autistic families, and are described elsewhere70. The AGRE cohort (n=482 unrelated autistic individuals) consists of simplex autism families, details of which are provided elsewhere71. In addition, we included 401 individuals (including 25 neurotypical individuals) from the EU-AIMS LEAP92 and Paris73 cohorts. Across all cohorts, we included only unrelated individuals, who were predominantly of European Ancestry as defined by genetic principal components (5 SD deviations above or below the mean of PC1 and PC2 from the HapMap CEU population). Additionally, we also included data from 1981 unrelated individuals (1000 males, 981 females) from the Nijmegen Biomedical Study (NBS) to provide support for the independent SNPs with $P < 1 \times 10^{-6}$ in the non-stratified GWAS. Participants were asked the question: If upsetts me if my daily routine is disturbed, which is related to a non-social domain of autism, and is similar to an item in the Autism Spectrum Quotient. Further information including genotyping and quality control is provided elsewhere83. Genetic association for the top SNPs were conducted using age, sex, and the first five genetic principal components as covariates using linear regression.

**Phenotypes.** The primary phenotype for this study is the SQ-R, which was used to conduct a GWAS in participants from 23andMe. The SQ-R is a self-report measure developed to specifically assess repetitive behaviour domain of the ADOS-G, a widely used instrument for diagnosing and assessing autism in four cohorts (SSC, AGRE, EU-AIMS LEAP, and Paris). Participants completed one of the following ADOS-G modules:1 1 (used for children with little or no phrase speech), 2 (for children with non-fluent speech), 3 (verbally fluent children), and 4 (verbally fluent adolescents and adults). For this study, we used the raw total of the scores from the social domain and the communication domain, combined. Scores for all four modules range from 0 to 24. The ADOS-G has high overall internal consistency, and high test–retest reliability for the social and communication subscales85. The choice for combining the social and communication domain scores were informed by factor analysis which suggested that the two domains contribute to one underlying factor85.

In contrast to the social and communication domain, the restricted and repetitive behaviour domain of the ADOS-G has poor test–retest reliability (r < 0.6) and a smaller range of scores (0–8) as it captures fewer repetitive and restrictive behaviour85. Hence, for this study, we used scores on the RBS86. The RBS-R is a measure developed to specifically measure restricted and repetitive behaviours in autistic individuals and captures stereotyped, self-injurious, sameness, compulsive, ritualistic, and restricted behaviour84, and has high inter-rater reliability and internal consistency85. The RBS-R comprises 45 questions with scores ranging from 3 to 1 for each item based on a Likert scale. ‘Insistence on sameness’ in the NBS cohort was measured using a single item: ‘It upsets me if my daily routine is disturbed’. This is related to a non-social domain of autism, and is again similar to an item in the Autism Spectrum Quotient. Participants were asked to indicate on a 4-point Likert scale ‘definitely agree’, ‘slightly agree’, ‘slightly disagree’, ‘definitely disagree’.

**Genotyping, imputation, and quality control and genetic association in the 23andMe cohort.** Details of genotyping, imputation and quality control in the 23andMe cohort are provided elsewhere47. Brieﬂy, unrelated participants were included if they had a call rate of >98.5%, and were of primarily European ancestry (97% European Ancestry). A total of 1,030,430 SNPs (including InDels) were used for imputation: SNPs were excluded if: they failed the Hardy–Weinberg equilibrium test at $P < 10^{-5}$, had a genotype rate of <90%; they failed the parent–offspring transmission test using trio data in the larger 23andMe research participant database; or if allele frequencies were signiﬁcantly different from the European 1000 Genomes reference data ($\chi^2$ test, $P < 10^{-5}$). Phasing was conducted using Beagle (version 3.3.1)93 in batches of 8000–9000 individuals. This was followed by imputation against all-ethnicity 1000 Genomes haplotypes (excluding monomorphic and singleton sites) using Minimac294. Genetic association analyses were restricted to SNPs with a minor-allele frequency > 1%. After quality control, 9,955,952 SNPs (imputed and genotyped) were included in the GWAS. Our primary analysis was of additive model of genetic effects and was conducted using a linear regression with age, sex, and the first five ancestry principal components included as covariates. In addition, given the modest sex difference, we also conducted sex-stratified analyses. SNPs were considered significant at a genome-wide threshold of $P < 5 \times 10^{-8}$. Leading SNPs were identified after LD-pruning using HapBank (r2 > 0.8). Winner’s curse correction was conducted using an FDR-based shrinking95.

We calculated variance explained by first standardising the regression estimates and then squaring the estimates. This is equivalent to: $R^2 = \frac{b^2}{\sigma^2}$, where $R^2$ is the proportion of variance explained for SNP, $b^2$ is the non-standardised regression coefficient, MAF is the minor-allele frequency for SNP, and $\sigma^2$ is the variance of SQ. Further details of this formula are provided in the Supplementary Note.

**Genotyping, imputation, and quality control and genetic association in the ALSPAC.** The SCDC91 scores were calculated from children of the 90 s (ALSPAC cohort), in children aged 8. In total, SCDC scores were available on N = 7,825 children. From this, we removed individuals for whom complete SCDC scores were not available. After excluding related individuals and individuals with high missingness (≥33%), and disproportionate heterozygosity. We restricted subsequent analyses to individuals of European descent (CEU), which were identified by multi-
dimensional scaling analysis and compared with Hapmap II (release 22).

Individuals were also removed if cryptic relatedness, assessed using identity by descent, was >0.1. Genotyped SNPs were filtered out if they had >5% missingness, violated Hardy–Weinberg equilibrium (P < 10−6), and had a minor allele frequency < 1%, resulting in a total of 526,688 genotyped SNPs. Haemotypes were estimated using data from mothers and children using ShapeIT (v2.6.44)88. Imputation was performed using Impute2 v2.2.290 against the 1000 genomes reference panel (release 1, Version 3). Imputed SNPs were excluded after all further analyses if they had a minor allele frequency < 1% and info < 0.8. After quality control, there were 8,282,911 genotyped and imputed SNPs that were included in subsequent analyses.

Dosage data from BGEN files were converted using hard-calls, with calls with uncertainty > 0.1 treated as missing data. Post-imputation, we excluded SNPs that deviated from Hardy–Weinberg equilibrium (P < 1 × 10−6), with minor allele frequency < 0.01 and missing call rates > 2%. We further excluded individuals with genotype missing rates > 5%. The SCDC score was normally distributed so we log-transformed the scores and ran regression analyses using the first two ancestry principal components and sex as the covariates using PLINK 2.0 (ref.90).

The log-transformed SCDC scores (henceforth, SCDC scores) had a modest but significant h2SNP as quantified using LDSR (h2SNP = 0.12 ± 0.05). LDSR intercept (0.99) suggested that there was no inflation in GWAS estimates due to population stratification. The λG was 1.013. We replicated the previously identified genetic correlation with autism57 (constrained intercept) using our SCDC GWAS (rG = 0.45 ± 0.18, P = 0.01). In addition, we also identified a negative genetic correlation between educational attainment58 and SCDC (rG = −0.30 ± 0.11, P = 0.007).

Genomic inflation factor, heritability, and functional enrichment for the SQ-R GWAS. LDSR91,92 was used to calculate for inflation in test statistics due to unaccounted population stratification. Heritability was calculated using LDSR using the first two West European LD scores. Difference in heritability between males and females was quantified using:

\[
Z = \frac{h^2_{\text{male}} - h^2_{\text{female}}}{\sqrt{SE_{\text{male}}^2 + SE_{\text{female}}^2}},
\]

where \(Z\) is the Z-score for the difference in heritability for a trait, \(h^2_{\text{male}}\) is the difference \(h^2_{\text{SNP}}\) estimate in males and females, and SE is the standard errors for heritability. Two-tailed P-values were calculated and reported as significant if \(P < 0.05\).

For the primary GWAS (non-stratified analyses), we conducted functional annotation using FUMA93. We restricted our analyses to the non-stratified analyses due to the high genetic correlation between the sexes and the low statistical power of the sex-stratified GWAS. We conducted gene-based association analyses using MAGMA94 within FUMA and report significant genes after using a stringent Bonferroni corrected \(P\)-value of 0.05. In addition, we conducted enrichment for tissue specific expression and pathway analyses within FUMA. For the significant SNPs, we investigated enrichment for eQTLs using brain tissues in the BRAINEAC and GTEx95 database within FUMA. We further conducted partitioned heritability for tissue-specific active chromatin marks and baseline functional categories using extended methods in LDSR96.

Hi-C-based annotations of fine mapped loci. We fine mapped three genome-wide significant loci (index SNPs: rs1446336 and rs1539586 or SQ; rs8005092 for SQ-R males) to obtain credible SNPs. First, we selected SNPs with \(P < 0.01\) that are located in the LD region (\(r^2 > 0.8\)) within the index SNP. LD structure within a locus was constructed by calculating correlations between SNPs within a locus (1KG v20130502). CAVIAR97 was then applied to the summary association statistics and a posterior probability of 0.95. In total, we identified 34 credible SNPs from the three GWAS loci.

For each locus, candidate genes were identified by mapping credible SNPs based on physical interactions in foetal brain as previously described98. One locus (index SNP rs1446336) was mapped to two genes, LSAMP and PTMAARP, indicating that two credible SNPs (rs13066948 and rs17173893) located in this locus physically interact with these genes.

Genetic correlation. For all phenotypes, we performed genetic correlation without constraining the intercept using LDSR. We identified significant genetic correlations using a Bonferroni adjusted \(P\)-value < 0.05. For the primary GWAS correlation analyses with SQ-R, we included psychiatric conditions57,63,99, personality traits100,101, measures of intelligence102,103, and social traits related to autism56,65,66, including scores on the SCDC, as previous research has investigated the phenotypic correlation between these domains and systemising104,105,106,107,108,109,110,111,112.

To understand the correlation between systemising and various phenotypes that have been genetically correlated with autism, we used GWAS data from 15 phenotypes including those that have a unique SQ-R phenotype after conditioning on the genetic effects of educational attainment110,111,112 to autism, we extended the beta coefficients, \(B\) is the regression coefficient obtained from the non-standardised GWAS, MAF is the minor allele frequency, \(\sigma^2\) is the variance of the SQ-R. This equation is explained in detail in the Supplementary Note. We conducted GWIS using only educational attainment as we are unclear whether the GWAS of cognitive aptitude56 was conducted on a standardised phenotype. Further, there is a high genetic correlation between cognitive aptitude and educational attainment. In addition to GWIS, to validate the findings, we conducted GSEM113,114, a complementary but independent method. GSEM uses the genetic correlations and covariances calculated using LDSR after accounting for sample overlap.

Polygenic scores in the SSC, AGRE, EU-AIMS LEAP, and Paris cohorts. We generated polygenic scores for SQ-R (mean weighted score of all the alleles that contribute to higher systemising) in 2221 probands from the SSC (Discovery dataset). We downloaded genotype data from the SSC from SFARI base (https://www.sfari.org/resource/sfari-base/). Individuals were genotyped on three different platforms: Illumina Omni2.5, Illumina 1Mv3, or Illumina 1Mv1. Informed consent or assent was obtained from all participants. In addition, the research team obtained ethical approval from the Cambridge Human Biology Research Ethics Committee to access and analyse the de-identified data from the SSC. We conducted stringent quality control and imputation separately for each platform. The full pipeline is available here: https://github.com/autism-research-centre/SSC-liftover-implementation. Briefly, individuals were excluded if they had: a genotyping rate < 95%, excessive or low heterozygosity (less or more than 3 SD from the mean), mismatched reported and genetic sex, and families with Mendelian errors > 3%. We further removed SNPs that significantly deviated from Hardy–Weinberg equilibrium (\(P < 1 \times 10^{-6}\)), had Mendelian errors > 10% of the families, and SNPs that were not genotyped > 10% of the families. We then conducted multi-dimensional scaling using the HapMap3 phase 3 population using the unrelated individuals CEU and TSI populations as representatives of the European population. This was conducted only in the parents to retain unrelated individuals for multi-dimensional scaling. Genetic principal components were calculated using only SNPs with minor allele frequency > 5%, and pruning the SNPs in Plink using an \(r^2\) of 0.2. We excluded families from further downstream analyses if either one of the parents were greater or less than 5 standard deviations from the means of the first two genetic principal components. Using only the unrelated individuals in HapMap3 CEU and TSI populations. Quality control was done using Plink v1.9. R. Phasing and imputation were conducted using the Michigan
Imputation Server (https://imputationserver.sph.umich.edu/start.html) using the 1000 genomes Phase 3 v5 as the reference panel. Polygenic scores were generated using PRSice2 (https://choishingwan.github.io/PRSice/) for the SQ-R using the non-stratified GWAS data. We calculated the mean polygenic score for each of the 2221 probands in the SSC, after clumping SNPs using an R² threshold of 0.1. Prior to generating polygenic scores, we confirmed that the probands were not related to each other using identity by descent PI-HAT > 0.15 as a relatedness cut-off. We used a P-value threshold of 1 as a previous research on educational attainment, subjective wellbeing and social relationship satisfaction, all suggest that the maximum variance explained is at a threshold of 1 (refs. 58,110). This is expected for highly polygenic traits where many SNPs incrementally contribute to the variance explained123. Polygenic scoring was done using standardised values on two different phenotypes as the dependent variable (RBS-R and the social and communication domain of the ADOS-G). We included sex, platform, the first 15 genetic principal components and standardised full-scale IQ as covariates. In addition, for the analysis of ADOS-G, we included the ADOS-G module as a covariate. Linear regression was conducted in R. A total of 135,233 SNPs were included in the polygenic score analyses after clumping and thresholding.

To validate the polygenic scores, we conducted additional polygenic score analysis using data combined from the AGRE, EU-AIMS LEAP and Paris cohorts. We followed similar quality control and imputation procedures to the SSC cohort. Given that this dataset was a mix of related and unrelated individuals, we chose unrelated individuals using a genomic relationship matrix (GRM) as provided in 1000 genomes Phase 3 v5 as the reference panel.

Data availability

The SQ-R GWAS results are available from 23andMe. The full set of summary statistics can be made available to qualified investigators who enter into an agreement with 23andMe that protects participant confidentiality. Interested investigators should email dataset-request@23andme.com for more information. Top SNPs (n = 10,000) can be visualised here: https://ghic Pasteur.fr/Data. For ALS PAC, it can be requested here: http://www.brightgenetics.co.uk/alspacs/permissions/access/. Data from the Simons Simplex Collection can be requested here: https://www.sfari.org/resource/sfari-base/. Summary GWAS statistics were downloaded from the PGC consortium: http://www.med.ucl.ac.uk/pgc/results-and-downloads/. Data for chromosome 12 was downloaded from http://www. t2d并与genes.org/data/. Data for self-reported tiredness was downloaded from http://www. cambridgesciencecut.ac.uk/node/335. Additional source data are available in Supplementary Data 1–13.

Code availability

Genomic-SEM: https://github.com/MichelNivard/GenomicSEM. GWIS: https://sites. google.com/site/mgnivid/GWIS. Plink: https://www.cog-genomics.org/plink2/ PRSice2: https://choishingwan.github.io/PRSice2/ CAVIAR: http://genetics.cs.ucla.edu/caviar/ Michigan Imputation Server: https://imputationserver.sph.umich.edu/index.html, custom code for quality control of the SSC and the other cohorts can be downloaded from https://github.com/autism-research-centre/SSC_liftover_imputation (https://doi. org/10.5281/zenodo.3342561) and from https://github.com/vwarrier/ PARIS_LEAP_analysis (https://doi.org/10.5281/zenodo.3342569).

Received: 20 December 2018 Accepted: 19 July 2019
Published online: 03 September 2019

References

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders 5th edn (American Psychiatric Publishing, Arlington, VA, 2013).
2. Happé, F. & Ronald, A. The ‘Fractionable Autism Triad’: a review of evidence from behavioural, genetic, cognitive and neural research. Neuropsychol. Rev. 18, 287–304 (2008).
3. Happé, F., Ronald, A. & Plomin, R. Time to give up on a single explanation for autism. Nat. Neurosci. 9, 1218–1220 (2006).
4. Shuster, J., Perry, A., Bebb, J. & Toplak, M. E. Review of factor analytic studies examining symptoms of autism spectrum disorders. J. Autism Dev. Disord. 44, 90–110 (2014).
5. Mandy, W. P. L. & Skuse, D. H. 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 (2008).
6. Palmer, C., Paton, B., Enticott, P. G. & Hohwy, J. Systemizing and empathizing in autism. J. ADHD 5, 797–805.e2 (2013).
7. Frazier, T. W., Youngstrom, E. A., Kubu, C. S., Sinclair, L. & Rezai, A. Exploratory and confirmatory factor analysis of the autism diagnostic interview-revised. J. Autism Dev. Disord. 38, 474–480 (2008).
8. Grove, R., Baillie, A., Allison, C., Baron-Cohen, S. & Hesketh, R. A. Empathizing, systemizing, and autistic traits: Latent structure in individuals with autism, their parents, and general population controls. J. Abnorm. Psychol. 122, 600–609 (2013).
9. Wheelwright, S. J. et al. Predicting autism spectrum quotient (AQ) from the systemizing quotient-revised (SQ-R) and empathy quotient (EQ). Brain Res. 1079, 47–56 (2006).
10. Svedholm-Häkkinen, A. M., Halme, S. & Lindeman, M. Empathizing and systemizing are differentially related to dimensions of autistic traits in the general population. Int. J. Clin. Heal. Psychol. 18, 35–42 (2018).
11. Greenberg, D. M., Warrier, V., Allison, C. & Baron-Cohen, S. Testing the empathizing-systemizing theory of sex differences and the extreme male brain hypothesis. J. Child Psychol. Psychiatry 45, 691–699 (2006).
12. Dvoryanski, K., Happé, F., Bolton, P. & Ronald, A. Relationship between symptom domains in autism spectrum disorders: a population based twin study. J. Autism Dev. Disord. 39, 1197–1210 (2009).
13. Norbury, C. F. Practitioner review: social (pragmatic) communication disorder conceptualization, evidence and clinical implications. J. Child Psychol. Psychiatry 55, 204–216 (2014).
14. Ulljarević, M., Evans, D. W., Alvares, G. A. & Whitehouse, A. J. O. Short repeat polymorphisms in the 5′ untranslated region of the 17q21.33: genes prioritized by expression in fetal brains. Mol. Psychiatry 19, 18–21 (2014).
15. Georgiades, S. et al. Investigating phenotypic heterogeneity in children with autism. J. Autism Dev. Disord. 43, 1079–1088 (2013).
16. Norbury, C. F. Practitioner review: social (pragmatic) communication disorder conceptualization, evidence and clinical implications. J. Child Psychol. Psychiatry 55, 204–216 (2014).
17. Ulljarević, M., Evans, D. W., Alvares, G. A. & Whitehouse, A. J. O. Short repeat polymorphisms in the 5′ untranslated region of the 17q21.33: genes prioritized by expression in fetal brains. Mol. Psychiatry 19, 18–21 (2014).
18. Georgiades, S. et al. Investigating phenotypic heterogeneity in children with autism spectrum disorder: a factor mixture modeling approach. J. Child Psychol. Psychiatry 54, 206–215 (2013).
19. St-Pourcin, B. et al. Variability in the common genetic architecture of social-communication spectrum phenotypes during childhood and adolescence. Mol. Autism 5, 18 (2014).
Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and... 4204–4207 (2014).

Wakabayashi, A. & Kawashima, H. Is empathizing in the E–S theory similar to agreeableness? The relationship between the EQ and SQ and major personality domains. Pers. Individ. Differ. 76, 88–93 (2015).

Deary, I. et al. Genetic contributions to self-reported tiredness. Mol. Psychiatry 23, 609–620 (2018).

Fuchsberger, C., Abecasis, G. R. & Hinds, D. A. minimac2: faster genotype imputation. Bioinformatics 31, 782–784 (2015).

Bigdelli, T. B. et al. A simple yet accurate correction for winner curse can predict signals discovered in much larger genome scans. Bioinformatics 32, 2598–2603 (2016).

De Manke, O., Marchini, J. & Zagury, J.-F. A linear complexity phased method for thousands of genomes. Nat. Methods 9, 179–181 (2011).

Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 5, e1000529 (2009).

Okbay, A. et al. Genetic variants associated with subjective well-being, personality, educational and occupational attainment. Nature 559, 89–94 (2018).

Russell-Smith, S. N., Bayliss, D. M., Maybery, M. T. & Tomkinson, R. L. Are the autism and positive schizotypy spectra diametrically opposed in empathizing and systemizing? J. Autism Dev. Disord. 43, 695–706 (2013).

Court, A. et al. Levels of autistic traits in anorexia nervosa: a comparative psychometric study. BMC Psychiatry 13, 222 (2013).

Nette, D. Empathizing and systemizing: what are they, and what do they contribute to our understanding of psychological sex differences? Br. J. Psychol. 99, 237–255 (2008).

Sims, S. H., J. W., Symons, F. J., Parker, D. E. & Lewis, M. H. Varieties of depression, and neuroticism identified in genome-wide association study of anorexia nervosa. J. Autism Dev. Disord. 43, 668–673 (2013).

Koo, F. M., Groen, Y., Becke, M., Fuernmaier, A. B. M. & Tucha, O. Self-reported empathy in adult women with autism spectrum disorders—a systematic mini review. PLoS ONE 11, e0151568 (2016).

Mattheisen, M. et al. Genome-wide association study in obsessive-compulsive disorder: results from the OCGAS. Mol. Psychiatry 20, 337–344 (2015).

International Obsessive Compulsive Disorder Foundation Genetics Collaborative (IOCDF-GC) and OCD Collaborative Genetics Association Studies (OCGAS), P. D. et al. Revealing the complex genetic architecture of obsessive-compulsive disorder using meta-analysis. Mol. Psychiatry 23, 1181–1188 (2018).

Duncan, L. E. et al. Largest GWAS of PTSD (N = 20,070) yields genetic overlap with schizophrenia and sex differences in heritability. Mol. Psychiatry 23, 666–673 (2017).

Bulik-Sullivan, B. K. et al. An atlas of genetic correlations across human diseases and traits. Nat. Genet. 47, 1236–1241 (2015).

Taitt, L. et al. Partitioning heritability by functional annotation using gene-set analysis of GWAS data. Am. J. Hum. Genet. 99, 917–927 (2016).

Nieto-Bohner, H. A., Pool, R., Dolan, C. V., Boomsma, D. I. & Nivard, M. G. GWIS: Genome-Wide Inferred Statistics for functions of multiple phenotypes. Am. J. Hum. Genet. 99, 917–927 (2016).

Allison, C., Baron-Cohen, S., Wheelwright, S. J., Stone, M. H. & Muncer, S. J. Psychometric analysis of the empathy quotient (EQ). Pers. Individ. Dif. 51, 829–835 (2011).

Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. Am. J. Hum. Genet. 88, 76–82 (2011).

Acknowledgements

We are grateful to Michel Nivard and Beate St. Pourcain for their help with the analytical methods. We thank the research participants and employees of 23andMe for making this study possible. We are grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, and nurses. V.W. was funded by St. John’s College, Cambridge, and the Cambridge Commonwealth Trust. This study was funded by grants from the National Institute for Health Research and Care East of England at Cambridgeshire and Peterborough Health Research and Care East of England at Cambridgeshire and Peterborough NHS Foundation Trust. We also received support from the NIHR Cambridge Biomedical Research Centre. We acknowledge with gratitude the generous support of Dr 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. Data obtained from 23andMe was supported by the National Human Genome Research Institute of the National Institutes of Health (grant number R44HG006981). The UK Medical Research Council and Wellcome (grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. GWAS data were generated by Sample Logistics and Genotyping Facilities at Wellcome Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe. This publication is the result of work from authors who will serve as guarantors for the content of this paper. The iPYSCH (The Lundbeck Foundation Initiative for Integrative Psychiatric Research) team acknowledges funding from The Lundbeck Foundation (grant nos. R102-A9118 and R155-2014-1724), the Stanley Medical Research Institute, the European Research Council
(project no.: 294838), the Novo Nordisk Foundation for supporting the Danish National Biobank resource, and grants from Aarhus and Copenhagen Universities and University Hospitals, including support to the iSEQ Center, the GenomeDK HPC facility, and the CIRRAU Center. The project leading to this application has 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. We thank the iPSYCH-Broad Autism Group and the EU-AIMS LEAP group for sharing data. A full list of the authors and affiliations in the iPSYCH-Broad autism group and the EU-AIMS LEAP group is provided in the Supplementary Information.

Author contributions
V.W., T.B., S.B.-C., and D.A.H. conceived and designed the analysis. C.S.L., F.C., R.D., W.D.W., J.B., A.D.B., J.G., G.P., 23andMe Research Team, and D.A.H. collected or contributed the data or analysis tools. V.W., R.T., H.W., F.C., and W.D.W. performed the analysis. V.W., R.T., B.C., D.A.H., T.B., and S.B.C. wrote the paper. D.A.H., T.B., and S.B.-C. supervised the analysis.

Additional information
Supplementary information accompanies this paper at https://doi.org/10.1038/s42003-019-0558-4.

Competing interests: D.A.H. and the 23andMe Research Team are employees of 23andMe, Inc. The remaining authors declare no competing interests.

The 23andMe Research Team
Michelle Agee12, Babak Alipanahi12, Adam Auton12, Robert K. Bell12, Katarzyna Bryc12, Sarah L. Elson12, Pierre Fontanillas12, Nicholas A. Furlotte12, Karen E. Huber12, Aaron Kleinman12, Nadia K. Litterman12, Jennifer C. McCreight12, Matthew H. McIntyre12, Joanna L. Mountain12, Carrie A.M. Northover12, Steven J. Pitts12, J. Fah Sathirapongsasuti12, Olga V. Sazonova12, Janie F. Shelton12, Suyash Shringarpure12, Chao Tian12, Joyce Y. Tung12, Vladimir Vacic12 & Catherine H. Wilson12