Systemizing is genetically correlated with autism and is

genetically distinct from social autistic traits

Varun Warrier[1,12], Roberto Toro[2,-4], Bhismadev Chakrabarti[1, 5], the iPSYCH-Broad autism group, Anders D Børglum[6-8], Jakob Grove[6-9], the 23andMe Research Team[10], David A. Hinds[10*,] Thomas Bourgeron[2,3,4*], and Simon Baron-Cohen[1,8,11,12*]

[1]Autism Research Centre, Department of Psychiatry, University of Cambridge, Cambridgeshire, United Kingdom

[2] Institut Pasteur, Human Genetics and Cognitive Functions Unit, Paris, France

[3] CNRS UMR 3571: Genes, Synapses and Cognition, Institut Pasteur, Paris, France

[4] Université Paris Diderot, Sorbonne Paris Cité, Human Genetics and Cognitive Functions, Paris, France

[5] Centre for Integrative Neuroscience and Neurodynamics, School of Psychology and Clinical Language Sciences, University of Reading, Reading, United Kingdom

[6] The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Denmark

[7] Centre for Integrative Sequencing, iSEQ, Aarhus University, Aarhus, Denmark

[8] Department of Biomedicine - Human Genetics, Aarhus University, Aarhus, Denmark

[9] Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark

[10] 23andMe Inc., Mountain View, California 94043, USA

[11] CLASS Clinic, Cambridgeshire and Peterborough NHS Foundation Trust (CPFT), Cambridgeshire, United Kingdom

[12] **Corresponding authors**: Varun Warrier (vw260@medschl.cam.ac.uk) and Simon Baron-Cohen (sb205@cam.ac.uk). Autism Research Centre, Department of Psychiatry, Cambridge University, Douglas House, 18B Trumpington Road, Cambridge, CB2 8AH, United Kingdom. Telephone: 0044 (0) 1223 746057. *Joint senior authors.

Abstract

The hypersystemizing theory of autism suggests that autistic individuals, on average, have superior attention to detail, and a stronger drive to systemize. Systemizing involves identifying *input-operation-output* relationships. Here, we report the results of genome-wide association studies (GWAS) of systemizing measured using the Systemizing Quotient - Revised in n = 51,564 individuals. We identify three genome-wide significant loci: Two of these were significant in the non-stratified GWAS: rs4146336 on chromosome 3 ($P = 2.58x10^{-8}$) and rs1559586 on chromosome 18 ($P = 4.78x10^{-8}$). In addition, we also identified a significant locus in the males-only GWAS (rs8005092 on chromosome 14, $P = 3.74x10^{-8}$). We find that 12%±1.2 of the variance in systemizing is captured by SNPs ($P=1.2x10^{-20}$). We identify a positive genetic correlation between autism and systemizing ($r_g = 0.26\pm0.06$; $P = 3.35x10^{-5}$), which is independent of genetic contribution to educational attainment. We further demonstrate that genetic risk for autism from systemizing is genetically distinct from genetic risk emerging from social autistic traits, suggesting distinct shared genetics between autism and social and non-social traits. Our results highlight the importance of considering both social and non-social autistic traits in elucidating the genetic architecture of autism.

Introduction

The hypersystemizing theory of autism suggests that autistic individuals, on average, have superior attention to detail, and a stronger drive to systemize. Systemizing involves identifying *input-operation-output* relationships in order to analyse and build systems and to understand the laws that govern specific systems[1]. Several lines of evidence suggest that autistic individuals have intact or superior systemizing. This idea was noted in the earliest reports describing autism. In his 1944 paper describing autism[2], Hans Asperger noted a proclivity for patterns and order in autistic children. Of one child, he writes, "He orders his facts into a system and forms his own theories even if they are occasionally abstruse." He observes that another child had "specialised technological interests and knew an incredible amount about complex machinery," while a third child "was preoccupied by numbers." In Leo Kanner's 1943 paper, he writes that the children with autism have "precise recollection of complex patterns and sequences[3]."

These initial clinical observations have been quantified using different measures. For example, on a self-report measure of systemizing (the Systemizing Quotient – Revised, or the SQ-R)[4], autistic adults, on average, score significantly higher than non-autistic individuals[4,5]. The same pattern of results is seen in autistic children, using the parent-report version of the SQ[6]. Systemizing is also highly correlated with aptitude in science, technology, engineering, and mathematics (STEM)[7]. Fathers and grandfathers of children with autism are significantly overrepresented in the field of engineering[8]. The same is true of mothers[9]. This is in line with higher rates of autism in geographical regions that have higher rates of people working in fields such as information technology, like Eindhoven in the Netherlands[10]. Further, autistic individuals are more likely to enrol in STEM majors (34.31%) compared to the general population (22.8%) and other learning disabilities (18.6%)[11]. STEM professionals also score significantly higher on measures of autistic traits (mean = 21.92, SD = 8.92) compared non-STEM professionals (mean = 18.92, SD = 8.48)[12]. Finally, unpublished work from Sweden suggests that high technical IQ in fathers increases risk for autism in children. A few studies have also investigated systemizing in other psychiatric traits and conditions, including schizotypy[13] and anorexia nervosa[14].

It is unclear if the link between autism and systemizing is due to underlying genetics or due to other non-biological factors (for example, people high in systemizing may be more aware of autism and hence more likely to seek a diagnosis). Here we directly test this using

genome-wide association data from 51,564 research participants from 23andMe, Inc., a personal genetics company, and who completed the SQ-R. This study has the following aims: 1. To investigate the genetic architecture of systemizing, using the SQ-R; 2. To identify the genetic correlation between the SQ-R and psychiatric conditions (including autism) and psychological traits; 3. To identify enrichment in tissues, gene groups, and biological pathways.

We identified three genome-wide significant loci associated with the SQ-R, and identified potential candidate genes for these loci using chromatin interactions and eQTLs in neural tissues. We identify a significant, and positive genetic correlation between the SQ-R and autism. This genetic correlation is independent of the genetic correlates of educational attainment, which is also genetically correlated with autism. We further demonstrate that social phenotypes which are genetically correlated with autism are not genetically correlated with the SQ-R, suggesting shared genetics between autism and two distinct phenotypic domains: social traits and non-social traits. In addition, the SQ-R is also significantly correlated with schizophrenia and measures of cognition. We also find that systemizing has a modest heritability, is enriched in evolutionary conserved sites and fetal DNAse hypersensitivity sites, and has a high genetic correlation between males and females.

Methods

Participants: Research participants in the GWAS of the SQ-R were from 23andMe and are described in detail elsewhere[15,16]. 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 questionnaire on the 23andMe participant portal. Only participants who were primarily of European ancestry (97% European Ancestry) were selected for the analysis using existing methods[17]. Unrelated individuals were selected using a segmental identity-by-descent algorithm[18]. A total of 51,564 participants completed the SQ-R (males = 26,063, and females = 25,501).

Phenotype: The SQ-R is self-report measure of systemizing drive, or interest in rule-based patterns[4]. There are 75 items on the SQ-R, with a maximum score of 150 and a minimum score of 0. Scores on the test are normally distributed. The SQ-R has good cross-cultural stability and good psychometric properties with Cronbach's alpha ranging from 0.79

to 0.94 in different studies[19]. Test-retest reliability available in a Dutch sample indicated high reliability of 0.79 (Pearson correlation)[19]. This was supported by another study in 4,058 individuals which identified high internal cohesion[20]. Exploratory followed by confirmatory factor analysis using Rasch modelling suggests that the SQ-R is unidimensional[20]. A sex difference has been observed in multiple studies with males, on average, scoring significantly higher than females. Criterion validity shows that the SQ-R has a modest but significant correlation with the Mental Rotation Test ($r = .25$, $P = .013$), as well as its subscales[21]. Autistic individuals, on average, score higher on the SQ-R in multiple different studies[22,23]. Further, the SQ-R also predicts autistic traits, with a combination of the SQ-R and the Empathy Quotient predicting as much as 75% of the variance on the Autism Spectrum Quotient, a measure of autistic traits[4]. In the Cambridge Autism Research Database, of 5,663 individuals, there were clear differences in scores on the SQ-R based on diagnosis, with autistic individuals scoring significantly higher on the SQ-R than non-autistic individuals, and the SQ-R being significantly and positively correlated with the AQ **(Supplementary Note)**.

Genotyping, imputation, and quality control: Details of genotyping, imputation and quality control are given elsewhere[24]. Briefly, unrelated participants were included if they had a call rate of greater than 98.5%, and were of primarily European ancestry (97% European ancestry). A total of 1,030,430 SNPs (including InDels) were genotyped. SNPs were excluded if: they failed the Hardy-Weinberg Equilibrium Test at $P < 10^{-20}$; had a genotype rate of less than 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-square test, $P < 10^{-20}$). Phasing was conducted using Beagle (version 3.3.1)[25] in batches of 8000-9000 individuals. This was followed by imputation against all-ethnicity 1000 Genomes haplotypes (excluding monomorphic and singleton sites) using Minimac2[26]. Genetic association analyses were restricted to SNPs with a minor allele frequency > 1%.

Genetic association: Our primary analysis was an additive model of genetic effects and was conducted using a linear regression with age, sex, and the first five ancestry principle 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 Plink ($r^2 > 0.8$). Winner's curse correction was conducted using an FDR based shrinking[27].

We calculate variance explained by first standardizing the regression estimates and then squaring the estimates. This is equivalent to:

$$R^2 = \hat{B}_j^2 \frac{(MAF_j)(1 - MAF_j)}{\sigma^2{}_y}$$

Where R^2 is the proportion of variance explained for SNP j. $\hat{B}_j^2$ is the non-standardized regression coefficient, MAF is the minor allele frequency for SNP j, and $\sigma^2{}_y$ is the variance of SQ. Further details of this formula are provided in the **Supplementary Note.**

GWAS of the SCDC: The SCDC (The Social and Communication Disorders Checklist)[28] scores were calculated from children of the 90s (ALSPAC cohort)[29], 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 no genetic data, data was available on a total of n = 5421 unrelated individuals. The SCDC score was not normally distributed so we log-transformed the scores and ran regression analyses using the first 2 genetic principal components and sex as the covariates using Plink 2.0[30]. Dosage data from BGEN files were converted using hard-calls, with calls with uncertainty > 0.1 treated as missing data. We excluded SNPs that deviated from Hardy-Weinberg Equilibrium ($P < 1x10^{-6}$), with minor allele frequency < 0.01 and missing call rates > 2%. We further excluded individuals with genotype missing rates > 5%. Further details are provided in the **Supplementary Note**.

Genomic inflation factor, heritability, and functional enrichment: LDSR[31,32] was used to calculate for inflation in test statistics due to unaccounted population stratification. Heritability and genetic correlation for all the traits was calculated using LDSR using the north-west European LD scores. For all phenotypes we performed genetic correlation without constraining the intercept. We identified significant genetic correlations using a Bonferroni adjusted P-value < 0.05. For autism, we validated the genetic correlation using a second, larger dataset: the PGC-iPSYCH meta-analysis[33] autism GWAS dataset **(Supplementary Note).** Details of the phenotypes, the sample sizes and the references are provided in **Supplementary Table 3.**

Heritability and genetic correlation was performed using extended methods in LDSR[32]. Difference in heritability between males and females was quantified using:

$$Z = \frac{h^2_{males} - h^2_{females}}{\sqrt{SE^2_{males} + SE^2_{females}}}$$

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

Functional annotations: For the primary GWAS (non-stratified analyses), we conducted functional annotation using FUMA[34]. 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 MAGMA[35] within FUMA and report significant genes after using a stringent Bonferroni corrected $P < 0.05$. 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 GTEx[36] database within FUMA. In addition, we also investigated chromatin interactions from the Dorsolateral Prefrontal Cortex, the hippocampus, and neural progenitor cells. Significant results were identified using a Benjamini-Hochberg FDR corrected $P < 0.05$. We further conducted partitioned heritability using extended methods in LDSR[37]. We conducted partitioned heritability for baseline categories.

GWIS: To investigate if the SQ-R is genetically correlated with autism independent of the genetic effects of cognition, we constructed a unique SQ-R phenotype after conditioning on the genetic effects of educational attainment using GWIS[38]. GWIS takes into account the genetic covariance between the two phenotypes to calculate the unique component of the phenotypes as a function of the genetic covariance and the SNP heritability. Prior to performing GWIS, we standardized the beta coefficients for the SQ-R GWAS by using the following formula:

$$\widehat{B_{std}} = \hat{B}\sqrt{\frac{2(MAF)(1 - MAF)}{\sigma^2{}_y}}$$

Where $\widehat{B_{std}}$ is the standardized regression coefficients, $\hat{B}$ is the regression coefficient obtained from the non-standardized GWAS, MAF is the minor allele frequency, $\sigma^2{}_y$ is the variance of the SQ-R. This equation is explained in detail in the **Supplementary Note**.

Phenotypic regression analyses: To identify if the SQ-R predicts attention to detail and rigidity as defined by Bralten and colleagues[39], we conducted regression analysis in n = 5,663 individuals from the Cambridge Autism Research Database (CARD) who had completed the AQ and the SQ-R. We derived the variables "attention to detail" and "rigidity" based on the scoring criteria provided by Bralten and colleagues[39]. We note that this is different from the attention to detail subscale as defined in the original paper describing the AQ[40]. Linear regression analyses were conducted using "attention to detail" and "rigidity" scores as the dependent variable, total SQ score as the predictor variable, and sex, status, and age of completion of both the SQ-R and the AQ as covariates. Status was divided into five categories: Autistic individuals, controls, first degree relative of an autistic individual, suspected autism, and suspected autism and relative of an autistic individual.

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 visualized here: https://ghfc.pasteur.fr

Results

Phenotypic distribution and genome-wide association analyses

The mean score for all participants was 71±21 out of a total of 150 on the SQ-R, with males scoring higher than females on the SQ-R (76.5±20 in males; 65.4±20.6 in females). The male-advantage was statistically significant ($P < 0.001$, Cohen's $d = 0.54$) **(Supplementary Figure 1)**.

Genome-wide association analyses identified three significant loci **(Figure 1)**. 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 ($P = 4.78 \times 10^{-8}$). In addition, we also identified a significant locus in the males-only GWAS (rs8005092 on chromosome 14, $P = 3.74 \times 10^{-8}$). LDSR intercept suggested that there was minimal inflation due to population stratification **(Figure 1)**. **Supplementary Table 1** provides a list of all independent SNPs with $P < 1 \times 10^{-6}$. For two of these loci (rs1559586 and rs8005092), eQTL and chromatin interactions identified 5 genes for the three loci: *TIMM21*, *FBXO15*, *RP11-169F17.1*, *NOVA1*, and *STXBP6*

(Supplementary Figure 2 and Supplementary Table 2). Of these, *STXBP6, NOVA1,* and *FBXO15* have relatively high expression in the brain in the GTEx database. *STXBP6* encodes a protein that binds to Syntaxin 1 and help assemble the SNARE complex, found in the presynaptic terminal[41]. *NOVA1* encodes a neuron-specific RNA binding protein that regulates neuronal miRNA[42]. *FBXO15* encodes an F-box domain containing protein which is involved in protein degradation via ubiquitination.

The most significant SNP in each GWAS analysis explained 0.001 – 0.0002% of the total variance (**Supplementary Table 3**) after accounting for winner's curse, suggesting considerable polygenicity and small per-SNP effect size. Local LD plots for the three significant loci are provided in **Supplementary Figure 3**.

insert Figure 1 here

Genetic correlation

We investigated genetic correlation between the SQ-R and psychiatric conditions, psychological traits, and measures of cognition **(Figure 2, Supplementary Table 4)**. The SQ-R was significantly and positively correlated with autism (PGC cohort) (r_g =0.22±0.06; P =1.3x10^{-3}) and schizophrenia (r_g = 14±0.04; P = 1.6x10^{-3}). We validated the association with autism using the larger PGC-iPSYCH autism cohort (r_g = 0.26±0.06; P = 3.35x10^{-5}). Among psychiatric conditions, the correlation between SQ and autism is the largest. In contrast, the correlation between SQ and schizophrenia is small. The SQ-R was also significantly and positively correlated with measures of cognition (years of schooling: r_g = 0.14±0.03, P = 4.73x10^{-05}, cognitive aptitude r_g = 0.19±0.04; P = 2.35x10^{-5}).

Systemizing is thought to contribute to the non-social domain of autism. To identify if systemizing contributes to social autistic traits we conducted genetic correlation using the Social and Communication Disorder Checklist (SCDC), which is genetically correlated with autism[43] **(Supplementary Note)**. Additionally, we also conducted genetic correlation between the SQ-R and family relationship satisfaction, friendship satisfaction[44], and self-report empathy which are negatively correlated with autism. We did not identify a significant genetic correlation between SQ-R and SCDC (r_g = 0.04±0.18, P = 0.81), family relationship satisfaction (r_g = -0.04 ± 0.06, P = 0.47), friendship satisfaction (r_g = 0.06±0.07, P = 0.42), and

empathy (r_g = 0.10±0.07, P = 0.17). Given that the sample size for the SCDC is small, we reran the genetic correlation after constraining the intercept, mirroring the analyses in the paper that reported a genetic correlation between autism and the SCDC. We did not identify significant genetic correlations even after constraining the intercept (r_g = 0.04±0.13, P = 0.73)) These results suggest that the contribution of systemizing to the genetic risk of autism is distinct from the contribution of social traits.

insert Figure 2 here

Educational attainment has a positive genetic correlation with both autism and SQ-R (**Supplementary Note**). To identify if the genetic correlation between autism and SQ-R is driven by educational attainment, we constructed an SQ-R GWAS conditioned on the genetic effects of educational attainment (SQminEdu GWAS) using GWIS (**Methods**). We confirmed that the SQminEdu GWAS was not genetically correlated with educational attainment (r_g = 0.063±0.032, P = 0.052), and was genetically correlated with SQ-R (r_g = 1.04±0.005 , P < $2.2x10^{-16}$). The SQminEdu GWAS was still genetically correlated with both the PGC autism dataset (r_g = 0.14±0.06, P = 0.032) and the PGC-iPSYCH autism dataset (r_g = 0.20±0.05, P = $4.6x10^{-4}$) **(Figure 3, Supplementary Table 5)**. These results suggest that SQ-R contributes to the risk for autism independent of the genetic effects of educational attainment.

insert Figure 3 here

A recent study[39] identified that polygenic scores for autism predicts autistic traits in specific subdomains of the Autism Spectrum Quotient (AQ) which they termed "attention to detail" and "rigidity." We queried if systemizing predicts both of these domains within the AQ. Whilst we did not have access to a cohort with both genetic data and AQ scores, we had information on n = 5,663 individuals who had completed both the AQ and the SQ-R in the Cambridge Autism Research Database. Our results show that the SQ-R significantly predicts variation in scores in the "attention to detail" and "rigidity" domains on the AQ as defined by Bralten et al., 2017[39] (Attention to detail: Beta = 0.025±0.0009, P < $2.2x10^{-6}$; Rigidity: Beta = 0.015±0.001, P < $2.2x10^{-6}$).

Heritability, enrichment analysis, and gene-based analyses

Additive SNP heritability calculated using LDSR was 0.12±0.012 for the SQ-R ($P=1.2x10^{-20}$). Further, there was no significant difference in additive SNP heritability between males and females (P = 0.34) **(Supplementary Figure 4 and Supplementary Table 6)**, which was strengthened by the high genetic correlation between males and females (1±0.17; $P = 3.91x10^{-10}$), suggesting a similar genetic architecture between sexes.Partitioned heritability identified enrichment in four functional categories for the SQ-R after FDR correction: evolutionarily conserved genetic regions in mammals (FDR P = 0.027), foetal DNAse hypersensitivity sites (FDR P = 0.039), histone H3K27 acetylation (FDR P = 0.039), and Hoffman transcription start sites (FDR P = 0.039) **(Supplementary Tables 7).** Gene based analysis conducted using MAGMA identified 4 significant genes below a Bonferroni-corrected threshold of $2.75x10^{-6}$: *SDCCAG8*, $P = 3.7x10^{-9}$; *ZSWIM6*, $P = 8.5x10^{-8}$; *ZNF574*, $P = 2x10^{-6}$; and *FUT8*, $P = 2.13x10^{-6}$ **(Supplementary Table 8)**. Pathway based and tissue enrichment analysis did not identify any significant results **(Supplementary Table 9 and Supplementary Figure 5)**.

Discussion

Here we present the results of a genome-wide association study of systemizing. We identify three genome-wide significant loci associated with SQ-R, two in the non-stratified GWAS and one in the males-only GWAS. These need to replicated in larger sample sizes to provide robust confirmation of the results. Chromatin interaction analyses in neural tissues prioritize five genes for further functional follow up. Three of these genes have high expression in brain tissues. One of these candidate genes is *STXBP6.* STXBP6 binds to STX1 in the presynaptic boutons to help inhibit Ca^{2+} mediated exocytosis of neurotransmitters[41]. Several studies have identified genes involved in synapse formation and maintenance that are disproportionately mutated in autism[45]. Further, transcriptome studies of the post-mortem brain have identified downregulated genes that are enriched for synaptic transmission[46,47].

The idea that autistic individuals, on average, have a stronger interest in and aptitude for systems is not new. Several theories have sought to understand this link. For example, the hyper-systemizing theory suggests that this could explain the ‘resistance to change’ seen in many autistic individuals, in that systemizing requires varying one variable at a time whilst holding all others constant, in order to observe how manipulating one variable (e.g, the operation) changes the output[1]. In this way, resistance to change may appear as a ‘symptom’

but may simply reflect a different learning style in which the autistic person is systematically seeking to understand a rule-governed domain. The prediction error theory may also be relevant to this hypothesis in that a strong systemizer may be more attracted to highly lawful, more predictable domains, like mathematics or computers, the social world being the extreme opposite of this[48]. Note that versions of this theory may not be well-supported by evidence, since this theory suggests autistic people will have difficulty making predictions across the board, whilst the hyper-systemizing theory suggests this will only be true in domains that are not lawful. While a few studies have investigated the genetic architecture of social traits, or repetitive behaviour, no study to our knowledge has investigated the genetic architecture of non-social cognitive traits associated with autism.

Genetic correlation analyses identified, as predicted, a positive correlation with autism. The correlation was not trivial, and this was independent of the genetic contributors to educational attainment. To further understand the genetic architecture of autism, we tested the genetic correlation between SQ-R and four social traits that are all correlated with autism: social and communication difficulties in childhood which is positively correlated with autism, and friendship satisfaction, family relationship satisfaction, and empathy, which are all negatively genetically correlated with autism. We did not identify a significant correlation between SQ-R and the four phenotypes. These results suggest that there are at least two distinct sources of shared genetics with autism: one linked to social autistic traits and another linked to non-social autistic traits. This lends itself to the idea of multiple underlying risk factors, genetic and other, that contribute to risk for autism. In this hypothesis, the risk for autism does not stem from one source of genetic risk, but multiple different sets of genetic risks that contribute to the considerable phenotypic heterogeneity observed within the autism spectrum. Further research is needed to understand if individuals within the autism spectrum can be stratified based on these sets of risk.

Approximately 12% of variance in systemizing was explained by SNPs, and this was similar between males and females. To our knowledge, there is no study examining heritability of the SQ-R in twins. Given the modest sex-difference observed in the phenotypic scores, we conducted sex-stratified GWAS. Genetic correlation analyses suggested a near identical genetic architecture between the sexes. This was also supported by a similar additive SNP heritability between the sexes. Taken together, these results suggest there is, by and large, a similar genetic architecture between sexes. The observed phenotypic sex differences

can then be attributed to differences in environment, or other biological factors (such as prenatal hormones) which we were unable to fully test in this study.

In conclusion, we conducted the first GWAS of systemizing and identify three genome-wide significant loci. Genetic correlation analyses identified significant and replicable positive correlation with autism. This is independent of educational attainment and is not correlated with social autistic traits. These results suggest that there are multiple different sources of genetic risk that contribute to autism.

References

1 Baron-Cohen S. The hyper-systemizing, assortative mating theory of autism. *Prog Neuropsychopharmacol Biol Psychiatry* 2006; **30**: 865–72.

2 Asperger H. 'Autistic psychopathy' in childhood. In: Frith U (ed). *Autism and Asperger syndrome*. Cambridge University Press: Cambridge, 1944, pp 37–92.

3 Kanner L. Autistic disturbances of affective contact. *Nerv Child J Psychopathol Psychother Ment Hyg Guid Child 2* 1943; : 217–50.

4 Wheelwright SJ, Baron-Cohen S, Goldenfeld N, Delaney J, Fine D, Smith R *et al.* Predicting Autism Spectrum Quotient (AQ) from the Systemizing Quotient-Revised (SQ-R) and Empathy Quotient (EQ). *Brain Res* 2006; **1079**: 47–56.

5 Baron-Cohen S, Richler J, Bisarya D, Gurunathan N, Wheelwright SJ. The systemizing quotient: an investigation of adults with Asperger syndrome or high-functioning autism, and normal sex differences. *Philos Trans R Soc Lond B Biol Sci* 2003; **358**: 361–74.

6 Auyeung B, Baron-Cohen S, Chapman E, Knickmeyer R, Taylor K, Hackett G. Foetal testosterone and the child systemizing quotient. *Eur J Endocrinol* 2006; **155**: S123–S130.

7 Nettle D. Empathizing and systemizing: what are they, and what do they contribute to our understanding of psychological sex differences? *Br J Psychol* 2007; **98**: 237–55.

8 Baron-Cohen S, Wheelwright SJ, Burtenshaw A, Hobson E. Mathematical talent is linked to autism. *Hum Nat* 2007; **18**: 125–131.

9 Windham GC, Fessel K, Grether JK. Autism spectrum disorders in relation to parental occupation in technical fields. *Autism Res* 2009; **2**: 183–191.

10 Roelfsema MT, Hoekstra RA, Allison C, Wheelwright SJ, Brayne C, Matthews FE *et al.* Are autism spectrum conditions more prevalent in an information-technology region? A school-based study of three regions in the Netherlands. *J Autism Dev Disord* 2012; **42**: 734–9.

11 Wei X, Yu JW, Shattuck P, McCracken M, Blackorby J. Science, technology, engineering, and mathematics (STEM) participation among college students with an autism spectrum disorder. *J Autism Dev Disord* 2013; **43**: 1539–46.

12 Ruzich E, Allison C, Chakrabarti B, Smith P, Musto H, Ring H *et al.* Sex and STEM occupation predict Autism-Spectrum Quotient (AQ) scores in half a million people. *PLoS One* 2015; **10**: e0141229.

13 Russell-Smith SN, Bayliss DM, Maybery MT, Tomkinson RL. Are the autism and positive schizotypy spectra diametrically opposed in empathizing and systemizing? *J Autism Dev Disord* 2013; **43**: 695–706.

14 Hambrook D, Tchanturia K, Schmidt U, Russell T, Treasure J. Empathy, systemizing, and autistic traits in anorexia nervosa: a pilot study. *Br J Clin Psychol* 2008; **47**: 335–9.

15 Tung JY, Do CB, Hinds DA, Kiefer AK, Macpherson JM, Chowdry AB *et al.* Efficient replication of over 180 genetic associations with self-reported medical data. *PLoS One* 2011; **6**: e23473.

16 Do CB, Tung JY, Dorfman E, Kiefer AK, Drabant EM, Francke U *et al.* Web-based genome-wide association study identifies two novel loci and a substantial genetic component for Parkinson's disease. *PLoS Genet* 2011; **7**: e1002141.

17 Eriksson N, Tung JY, Kiefer AK, Hinds DA, Francke U, Mountain JL *et al.* Novel associations for hypothyroidism include known autoimmune risk loci. *PLoS One* 2012; **7**: e34442.

18 Henn BM, Hon LS, Macpherson JM, Eriksson N, Saxonov S, Pe'er I *et al.* Cryptic distant relatives are common in both isolated and cosmopolitan genetic samples. *PLoS One* 2012; **7**: e34267.

19 Groen Y, Fuermaier ABM, Den Heijer AE, Tucha O, Althaus M. The Empathy and Systemizing Quotient: The psychometric properties of the Dutch version and a review of the cross-cultural stability. *J Autism Dev Disord* 2015; **45**: 2848–64.

20 Allison C, Baron-Cohen S, Stone MH, Muncer SJ. Rasch modeling and confirmatory factor analysis of the systemizing quotient-revised (SQ-R) scale. *Span J Psychol* 2015; **18**: E16.

21 Ling J, Burton TC, Salt JL, Muncer SJ. Psychometric analysis of the systemizing quotient (SQ) scale. *Br J Psychol* 2009; **100**: 539–552.

22 Baron-Cohen S, Cassidy S, Auyeung B, Allison C, Achoukhi M, Robertson S *et al.* Attenuation of typical sex differences in 800 adults with autism vs. 3,900 controls. *PLoS One* 2014; **9**: e102251.

23 Wakabayashi A, Baron-Cohen S, Uchiyama T, Yoshida Y, Kuroda M, Wheelwright S. Empathizing and systemizing in adults with and without autism spectrum conditions: Cross-cultural stability. *J Autism Dev Disord* 2007; **37**: 1823–1832.

24 Warrier V, Grasby K, Uzefovsky F, Toro R, Smith P, Chakrabarti B *et al.* Genome-wide meta-analysis of cognitive empathy: heritability, and correlates with sex, neuropsychiatric conditions and cognition. *Mol Psychiatry* 2017.

25 Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies By use of localized haplotype clustering. *Am J Hum Genet* 2007; **81**: 1084–1097.

26 Fuchsberger C, Abecasis GR, Hinds DA. minimac2: faster genotype imputation. *Bioinformatics* 2015; **31**: 782–784.

27 Bigdeli TB, Lee D, Webb BT, Riley BP, Vladimirov VI, Fanous AH *et al.* A simple yet accurate correction for winner's curse can predict signals discovered in much larger genome scans. *Bioinformatics* 2016; **32**: 2598–2603.

28 Skuse DH, Mandy WPL, Scourfield J. Measuring autistic traits: heritability, reliability and validity of the Social and Communication Disorders Checklist. *Br J Psychiatry* 2005; **187**: 568–572.

29 Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson J *et al.* Cohort Profile: The 'Children of the 90s'—the index offspring of the Avon Longitudinal

Study of Parents and Children. *Int J Epidemiol* 2013; **42**: 111–127.

30 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559–75.

31 Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Patterson N *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* 2015; **47**: 291–295.

32 Bulik-Sullivan BK, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* 2015; **47**: 1236–41.

33 Pedersen CB, Bybjerg-Grauholm J, Pedersen MG, Grove J, Agerbo E, Bækved-Hansen M *et al.* The iPSYCH2012 case–cohort sample: new directions for unravelling genetic and environmental architectures of severe mental disorders. *Mol Psychiatry* 2017. doi:10.1038/mp.2017.196.

34 Watanabe K, Taskesen E, van Bochoven A, Posthuma D. FUMA: Functional mapping and annotation of genetic associations. *bioRxiv* 2017.

35 de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* 2015; **11**: 1–19.

36 Ardlie KG, Deluca DS, Segre A V., Sullivan TJ, Young TR, Gelfand ET *et al.* The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science (80-)* 2015; **348**: 648–60.

37 Finucane HK, Bulik-Sullivan BK, Gusev A, Trynka G, Reshef Y, Loh P-R *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* 2015; **47**: 1228–1235.

38 Nieuwboer HA, Pool R, Dolan C V., Boomsma DI, Nivard MG. GWIS: Genome-Wide Inferred Statistics for functions of multiple phenotypes. *Am J Hum Genet* 2016; **99**: 917–927.

39 Bralten J, van Hulzen KJ, Martens MB, Galesloot TE, Arias Vasquez A, Kiemeney LALM *et al.* Autism spectrum disorders and autistic traits share genetics and biology. *Mol Psychiatry* 2017. doi:10.1038/mp.2017.98.

40 Baron-Cohen S, Wheelwright SJ, Skinner R, Martin J, Clubley E. The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *J Autism Dev Disord* 2001; **31**: 5–17.

41 Constable JRL, Graham ME, Morgan A, Burgoyne RD. Amisyn regulates exocytosis and fusion pore stability by both Syntaxin-dependent and Syntaxin-independent mechanisms. *J Biol Chem* 2005; **280**: 31615–31623.

42 Storchel PH, Thummler J, Siegel G, Aksoy-Aksel A, Zampa F, Sumer S *et al.* A large-scale functional screen identifies Nova1 and Ncoa3 as regulators of neuronal miRNA function. *EMBO J* 2015; **34**: 2237–2254.

43 Robinson EB, St Pourcain B, Anttila V, Kosmicki JA, Bulik-Sullivan BK, Grove J *et al.* Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. *Nat Genet* 2016; **48**: 552–5.

44 Warrier V, the 23andMe Research team, Bourgeron T, Baron-Cohen S. Genome-wide association study of social relationship satisfaction: significant loci and correlations with psychiatric conditions. *bioRxiv* 2017; : 196071.

45 Bourgeron T. From the genetic architecture to synaptic plasticity in autism spectrum disorder. *Nat Rev Neurosci* 2015; **16**: 551–563.

46 Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, Horvath S *et al.* Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* 2011; **474**: 380–384.

47 Parikshak NN, Swarup V, Belgard TG, Irimia M, Ramaswami G, Gandal MJ *et al.* Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. *Nature* 2016; **540**: 423–427.

48 Sinha P, Kjelgaard MM, Gandhi TK, Tsourides K, Cardinaux AL, Pantazis D *et al.* Autism as a disorder of prediction. *Proc Natl Acad Sci U S A* 2014; **111**: 15220–5.

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 work possible. We specifically thank the following members of the 23andMe Research Team: Michelle Agee, Babak Alipanahi, Adam Auton, Robert K. Bell, Katarzyna Bryc, Sarah L. Elson, Pierre Fontanillas, Nicholas A. Furlotte, Karen E. Huber, Aaron Kleinman, Nadia K. Litterman, Jennifer C. McCreight, Matthew H. McIntyre, Joanna L. Mountain, Carrie A.M. Northover, Steven J. Pitts, J. Fah Sathirapongsasuti, Olga V. Sazonova, Janie F. Shelton, Suyash Shringarpure, Chao Tian, Joyce Y. Tung, Vladimir Vacic, and Catherine H. Wilson. We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, and nurses. VW was funded by St. John's College, Cambridge, and the Cambridge Commonwealth Trust. This study was funded by grants from the Medical Research Council, the Wellcome Trust, the Autism Research Trust, the Templeton World Charity Foundation, the Institut Pasteur, the CNRS and the University Paris Diderot. The research was conducted in association with the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care East of England at Cambridgeshire and Peterborough NHS Foundation Trust. We acknowledge with gratitude the generous support of Drs Dennis and Mireille Gillings in strengthening the collaboration between SBC and TB, 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 was 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 work of the authors who will serve as guarantors for the content of this paper. The iPSYCH (The Lundbeck Foundation Initiative for Integrative Psychiatric Research) team acknowledges funding from The Lundbeck Foundation (grant no 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.

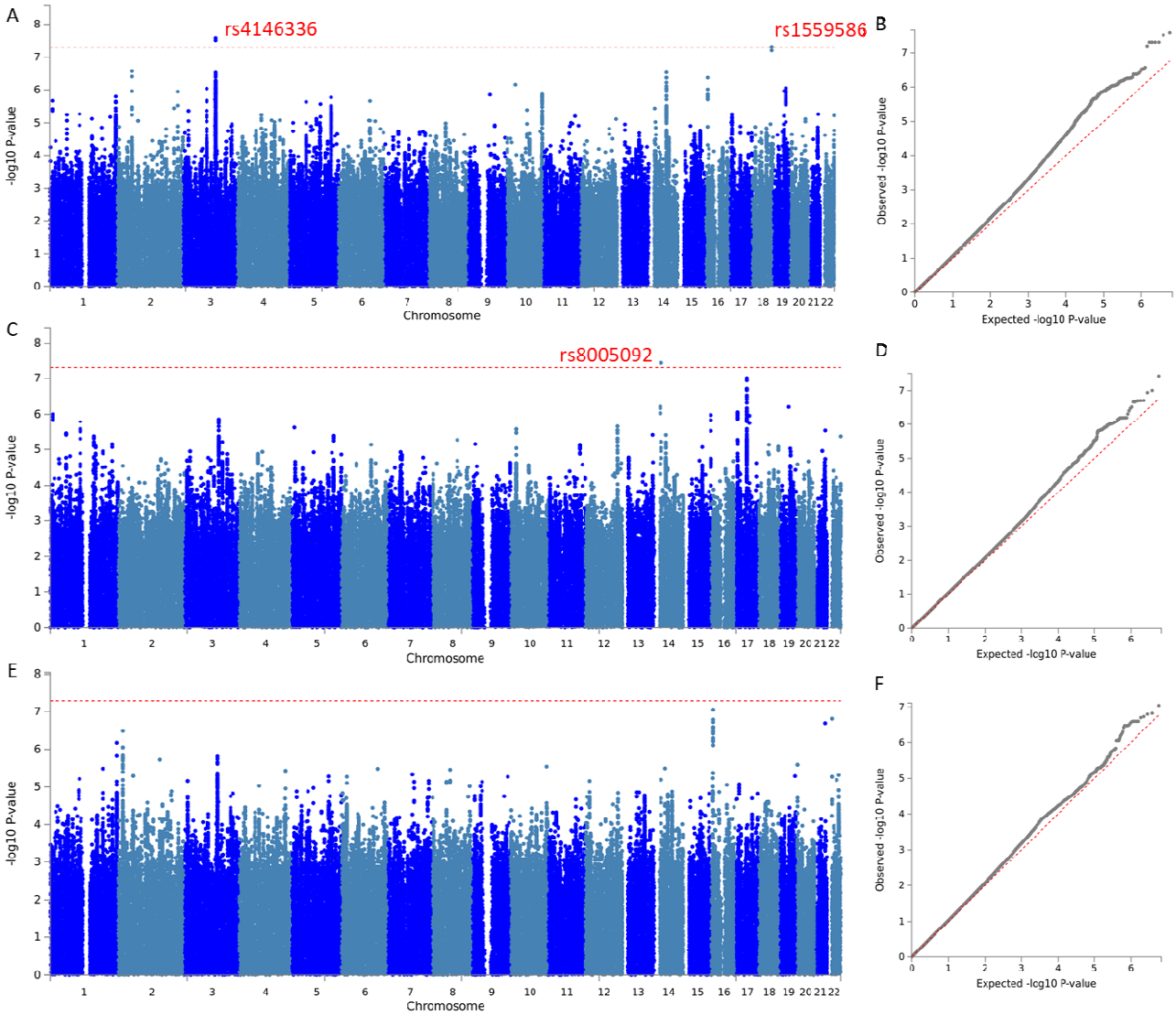
A full list of the authors and affiliations in the iPSYCH-Broad autism group is provided in the Supplementary Information.

Conflict of Interest

DH and the 23andMe Research Team are employees of 23andMe, Inc.

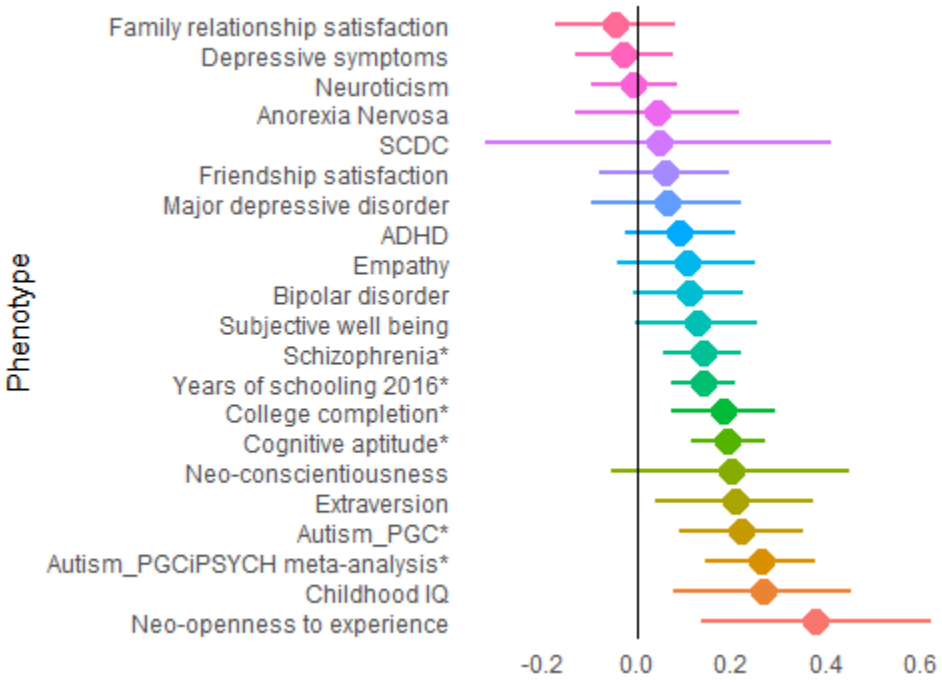
Figures

Figure 1: Manhattan and QQ-plots for the three GWAS



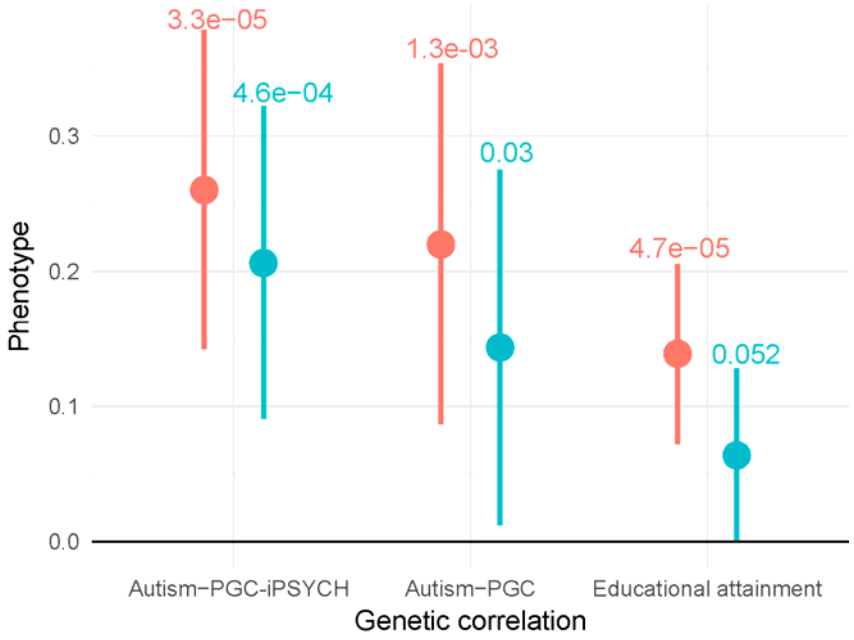
Manhattan plot for the three SQ-R GWAS: non-stratified (A), males-only (C), females-only (E). Significant SNPs are highlighted in red. QQ-plots for the three SQ-R GWAS: non-stratified (B), males-only (D), females-only (F). SQ-R nonstratified (N = 51,564): λ_{GC} = 1.10, LDSR intercept = 0.99, SQ-R males only (N = 26,063): λ_{GC} = 1.06, LDSR intercept = 0.99, SQ-R females only (N =25,051): λ_{GC} = 1.05, LDSR intercept = 1.01.

Figure 2: Genetic correlation between the SQ-R and other phenotypes



*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 Table 4. The following genetic correlations were significant after Bonferroni correction (indicated with * in the Figure): Autism PGC (r_g =0.22±0.06; P =1.3x10^{-3}, N = 16,350) Autism PGCiPSYCH (r_g = 0.26±0.06; P = 3.35x10^{-5}, N = 46,350), Years of Schooling 2016 (r_g = 0.13±0.03; P = 4.73x10^{-5}, N = 293723), College completion (r_g = 0.18±0.05; P = 1.30x10^{-3}, N = 95427), Cognitive aptitude (r_g = 0.19±0.04; P = 2.35x10^{-5}, N = 78,308), and Schizophrenia (r_g = 0.13±0.04; P = 1.6x10^{-3}, N = 77,096).*

Figure 3: Genetic correlations between SQ-R and SQminEdu GWIS estimates



Genetic correlation between SQ-R and three phenotypes (in Red). Genetic correlation between the GWIS SQminEdu dataset and the same three phenotypes (in blue). The bars represent 95% confidence intervals. P-values provided on top of the bars.