bioRxiv preprint doi: https://doi.org/10.1101/050682; this version posted February 23, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

# Genome-wide analyses of empathy and systemizing: correlations with psychiatric conditions, psychological traits, and education

**Short title: GWAS of empathy and systemizing**

Varun Warrier, MSc[1,8], Roberto Toro, PhD[2,3,4], Bhismadev Chakrabarti, PhD[1, 5],
Nadia Litterman, PhD[6], David A. Hinds, PhD[6,]
Thomas Bourgeron, PhD[2,3,4*], and Simon Baron-Cohen, PhD[1,7*]

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

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

[8] **Corresponding author**: Varun Warrier (vw260@medschl.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: 212 words. Article: 3,537 words. Tables: 1. Figures: 2. Supplementary material: 3

bioRxiv preprint doi: https://doi.org/10.1101/050682; this version posted February 23, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.


## Abstract

Background: Empathy and systemizing are dimensions of individual differences in cognition, and reveal areas of difficulty or strength in different psychiatric conditions. However, it is unclear how they contribute to genetic risk or resilience in various psychiatric conditions.

Methods: Here we report the largest genome-wide association study (GWAS) to date of empathy and systemizing using the Empathy Quotient (EQ) (n = 46,861) and the Systemizing Quotient-Revised (SQ-R) (n = 51,564) in participants from 23andMe, Inc.

Results: Both traits had a modest but significant narrow-sense heritability (EQ: $0.11 \pm 0.014$; $P= 1.7x10^{-14}$ and SQ-R: $0.12 \pm 0.012$; $P=1.2x10^{-20}$). As predicted, based on earlier work, we confirmed a female-advantage on the EQ and a male-advantage on the SQ-R in our large sample. However, we did not identify significant differences in heritability between sexes. We found significant genetic correlation between the variants conferring higher EQ scores and higher risk for schizophrenia ($P_{FDRadjusted} = 1.84x10^{-4}$) and anorexia nervosa ($P_{FDRadjusted} = 4.05x10^{-3}$), extraversion, NEO-conscientiousness, and subjective wellbeing. The genetic variants conferring higher scores for the SQ-R were correlated with higher risk for schizophrenia ($P_{FDRadjusted} = 8.64x10^{-3}$), with NEO-Openness to experience, and measures of educational attainment.

Conclusions: This study provides the first genetic confirmation of the roles of empathy and systemizing in psychiatric conditions, psychological traits and education.


## Introduction

Empathy is the ability to identify other people's thoughts, intentions, desires, and feelings, and to respond to other's mental states with an appropriate emotion (1). It plays an important role in social interaction and is a key component of both prosocial behaviour and social cognition. Differences in empathy have been found in several psychiatric conditions, including autism (1), bipolar disorder (2), schizophrenia (3–5), and major depressive disorder (2, 6, 7). Empathy is modestly heritable (8–10), and a few candidate gene association studies have investigated the role of various genes in empathy (11–13).

Systemizing is the drive to identify patterns to understand and build rule-based systems (14). Elevated systemizing has been identified in autism (14, 15). Individuals with autism, on average, tend to score higher than typical controls on tests of systemizing,

bioRxiv preprint doi: https://doi.org/10.1101/050682; this version posted February 23, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

including the Systemizing Quotient-Revised (SQ-R)(15). A few studies have investigated systemizing in other psychiatric traits and conditions, including schizotypy (16) and anorexia nervosa (17).

Because empathy and systemizing contribute to strengths and difficulties in several psychiatric conditions, they are important phenotypes for investigation. Understanding the biological networks that underlie these traits may help us understand how they contribute to psychiatric phenotypes, an approach that has been used for other traits such as neuroticism (18) creativity (19), and cognitive ability (20). Both empathy and systemizing show marked sex differences, in opposite directions: there is a male advantage in systemizing (15) and a female advantage in empathy (1). These sex differences are thought to contribute to the high proportion of males diagnosed with autism, although this could also in part reflect diagnostic practice (21). The sex difference in empathy and systemizing may also relate to the higher proportion of males in science, technology, engineering and mathematics (STEM) (which are inanimate object-centred or abstract sciences) and the higher proportion of females in medicine, veterinary sciences and psychology (which are animate- or people-centred sciences) (22). Both traits are correlated in opposite directions with levels of prenatal testosterone (23, 24), which itself is produced in higher levels in males, and which changes brain development and gene expression.

In this study, we investigate the genetic correlates of empathy and systemizing and how they contribute to risk for various psychiatric conditions. We performed sex-stratified and non-stratified genome-wide association analyses of empathy and systemizing in 23andMe, a personalized genetics company. We calculated the narrow sense heritability explained by all the SNPs tested, and investigated sex differences. We conducted genetic correlation analyses with five psychiatric conditions, six psychological traits, and measures of educational attainment.

## Results

### *Genome-wide association analyses*

To understand the genetic architecture of empathy and systemizing, we collaborated with 23andMe to conduct a Genome Wide Association Study (GWAS) of empathy (n = 46,861) and systemizing (n = 51,564). A flow chart of the study protocol is provided in Supplementary Figure 1. We employed two widely used self-report measures to quantify the

bioRxiv preprint doi: https://doi.org/10.1101/050682; this version posted February 23, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

traits: the Empathy Quotient (EQ) for empathy (1) and the revised Systemizing Quotient (SQ-R) for systemizing (15). The mean score for all participants was 46.4±13.7 on a total of 80 on the EQ, and 71±21 on a total of 150 on the SQ-R. Males scored higher than females on the SQ-R (76.5±20 in males; 65.4±20.6 in females), and vice-versa on the EQ (41.9±13.5 in males, 50.4±12.6) (Figure 1a and b). For each trait, we conducted three GWAS analyses: a male-only analysis, a female-only analysis, and a non-stratified analysis, using a linear regression model with age, and the first four ancestry principal components as covariates (Online Methods). We corrected for the three different tests for each trait and used a conservative threshold of $P = 1.66x10^{-8}$. We did not identify any genome wide significant SNPs (Supplementary Figures 2 - 17 and Supplementary Table 1). The most significant SNP in each GWAS analysis explained 0.06 – 0.13% of the total variance (Supplementary Information and Supplementary Table 19), although this will be lower after accounting for the 'winner's curse' or regression to the mean in effect estimate.

*Gene-based association and heritability*

Gene based analysis identified one significant putative gene, CMT1A Duplicated Region Transcript 4 (*CDRT4)* ($P_{FDRadjusted} = 8x10^{-4}$) for empathy (Supplementary Tables 2 and 3). We identified enrichment in four functional categories for the SQ-R after FDR correction. These are evolutionarily conserved genetic regions in mammals (P = 0.0005), fetal open chromatin sites (P = 0.002), histone acetylation sites (P = 0.002), and transcription start sites (P = 0.002) (Supplementary Tables 4 and 5). We used LDSC (25) to calculate the heritability explained by all the SNPs tested (Online Methods). The heritability was 0.11±0.014 for the EQ ($P= 1.7x10^{-14}$), and 0.12±0.012 for the SQ-R ($P=1.2x10^{-20}$; Figure 1*c* and *d* and Supplementary Table 6).

(insert Figure 1 here)

*Sex differences*

Sex differences in empathy and systemizing (15, 21) may reflect genetic as well as non-genetic factors (such as prenatal steroid hormones, and postnatal learning) (26). In our dataset, there was significant female advantage on the EQ ($P < 2x10^{-16}$ Cohen's d = 0.65), and a significant male advantage on the SQ-R ($P < 2x10^{-16}$; Cohen's d = 0.54) (Figure 1a and b). To investigate the biological basis for the sex-difference observed in the traits, we

bioRxiv preprint doi: https://doi.org/10.1101/050682; this version posted February 23, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

investigated the heritability of the sex-stratified GWAS analyses for both the traits. Our analyses revealed no significance difference between the heritability in the males-only and the females-only datasets for the two traits (P = 0.48 for male-female difference in EQ, and P = 0.34 for male-female difference in SQ-R) (Figure 1c and d; Supplementary Table 6). Additionally, there was a high genetic-correlation between the males-only and females-only GWAS for both the traits (EQ correlation = $0.82 \pm 0.16$; P = $2.34 \times 10^{-7}$; SQ-R correlation = $1 \pm 0.17$; P = $3.91 \times 10^{-10}$), indicating a high degree of similarity in the genetic architecture of the traits in males and females. However, there was significant heterogeneity in the effect estimates of the top SNPs ($P < 1 \times 10^{-6}$). (See supplementary Figure 18 and 19 and Supplementary Table 9).

*Genetic correlations*

To investigate how the two traits correlate with psychiatric conditions, psychological traits and educational attainment, we performed genetic correlation (Online Methods) with five psychiatric conditions (autism, ADHD, anorexia nervosa, bipolar disorder, depression (major depressive disorder and the larger depressive symptoms dataset) and schizophrenia), six psychological traits (NEO-extraversion, NEO-openness to experience, NEO-conscientiousness, neuroticism, and subjective wellbeing) and educational attainment (a proxy measure of IQ, measured using years of schooling) (Supplementary Table 7). With psychiatric conditions, three correlations were significant after correcting for multiple comparisons: EQ-schizophrenia ($r_g$ = 0.19±0.04; $P_{_FDRadjusted}$ = $1.84 \times 10^{-4}$), SQ-R-schizophrenia ($r_g$ = 14±0.04; $P_{_FDRadjusted}$ = $8.64 \times 10^{-3}$) and EQ-anorexia nervosa ($r_g$ = 0.32±0.09; $P_{_FDRadjusted}$ = $4.05 \times 10^{-3}$). As anorexia nervosa is primarily diagnosed in women, we wanted to test if the correlation is driven by the presence of higher female cases in the anorexia nervosa GWAS (68% females and 32% males) that could simply reflect higher EQ in the anorexia dataset rather than pleiotropy. We conducted correlation analyses using the females-only EQ (EQ-F) dataset and the anorexia nervosa dataset. The EQ-F-anorexia correlation was highly significant ($r_g$ = 0.48±0.12; P = $8.46 \times 10^{-5}$), confirming the underlying pleiotropy between empathy and anorexia nervosa. The EQ-M anorexia correlation was considerably lower but significant ($r_g$ = 0.15±0.11; P = 0.17).

With psychological traits, we identified four significant correlations: EQ with extraversion ($r_g$ = 0.45±0.08; $P_{_FDRadjusted}$ = $1.56 \times 10^{-06}$), subjective wellbeing ($r_g$ = 0.19±0.07;

bioRxiv preprint doi: https://doi.org/10.1101/050682; this version posted February 23, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$P_{_FDRadjusted}$ $2.97x10^{-2}$,), and NEO-conscientiousness ($r_g$ = 0.39±0.14; $P_{_FDRadjusted}$ = $2.97x10^{-02}$), and SQ-R with NEO-Openness to experience ($r_g$ = 0.37±0.12; $P_{_FDRadjusted}$ = $1.17x10^{-2}$). We also identified a significant association between SQ-R and years of schooling ($r_g$ = 0.14±0.03; $P_{_FDRadjusted}$ = $4.26x10^{-04}$), which we replicated using other measures of IQ such as childhood IQ ($r_g$ = 0.26±0.09, P = 0.0056) and college completion ($r_g$ = 0.18±0.056, P = 0.0013). (Figure 2; Supplementary Table 7). The correlation between the EQ and the SQ-R was non-significant ($r_g$ = 0.10±0.07; $P_{_FDRadjusted}$ = 0.36).

(insert Figure 2 here)

**Discussion**

This is the first GWAS to investigate the genetic architecture of empathy and systemizing. We identified several significant genetic correlations with EQ and SQ-R and psychiatric conditions, psychological traits, and education, providing insights into the shared genetic architecture. Although we did not identify any significant SNPs after correcting for multiple testing, we identified four SNPs with $P < 5x10^{-8}$. Males and females perform differently on the tests, and increasing the sample size is likely to detect SNPs that contribute in a sex-specific manner to both the traits.

We identified significant positive genetic correlations for EQ with schizophrenia and anorexia nervosa, and for SQ-R with schizophrenia. The empirical literature about empathy and schizophrenia in general report deficits in cognitive empathy (3, 27), but preserved or stronger affective empathy (5, 27) and emotional contagion (5) in individuals with schizophrenia compared to controls. The EQ measures both affective and cognitive components of empathy, and our results suggest that a subset of genetic variants associated with empathy also increase the risk for schizophrenia and anorexia nervosa; the latter remained significant after using the females-only EQ dataset. It is also worth noting that schizophrenia and anorexia share significant positive genetic correlation (rg = 0.23±0.06), and it is possible that the pleiotropy between these two conditions may be mediated by genetic variants that contribute to empathy. A study from our group also identified a significant genetic correlation between cognitive empathy and anorexia nervosa, underscoring the importance of empathy as a genetic risk factor in anorexia nervosa (28). A few studies have identified a correlation between empathy and psychosis (29, 30). To our knowledge, no study has identified a correlation between systemizing and schizophrenia. LDHub reports a positive genetic overlap with educational attainment (years of schooling),

bioRxiv preprint doi: https://doi.org/10.1101/050682; this version posted February 23, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

but a recent study using measures of cognitive aptitude in the UKBiobank identifies a negative, significant correlation between schizophrenia and general cognitive aptitude (31). Investigating using other measures of systemizing such as STEM aptitude will help validate the results.

Investigating genetic correlation results with psychological traits and measures of cognition further helped elucidate the genetic architecture of the two traits. The EQ was significantly correlated with subjective wellbeing, extraversion and conscientiousness. Both extraversion and conscientiousness contribute to empathy (32) which, in turn, contributes to subjective wellbeing (33). The SQ-R, on the other hand, had significant positive genetic correlations with different measures of cognition and NEO-openness to experience, which are genetically ($r_g$ = 0.40± 0.094; P = $1.79x10^{-05}$) (31, 34) and phenotypically correlated with each other (35, 36). Our research suggests a common underlying genetic architecture that contributes to the correlation between the three traits. Interestingly, both NEO-Openness to experience and educational attainment ($r_g$ = 0.342±0.06 P = $2.49x10^{-07}$) (34) are correlated with autism (rg = 0.42±0.13, P = 0.002) (34). The positive correlation between SQ-R and autism was not significant perhaps due to being underpowered..

Both the EQ and the SQ-R are self-report and it is unclear how much of the intrinsic biological variation in these traits is captured by the two tests. The EQ has three primary underlying factors (cognitive empathy, emotional reactivity and social skills) (37), whereas the SQ-R has one underlying factor (38). In this study, we were unable to stratify based on each factor and heritabilities and genetic correlations are for the broad trait rather than for specific factors. These subtleties must be kept in mind when interpreting the results.

This is also the first study to provide estimates of additive heritability explained by all the SNPs tested for the two traits. Approximately 11% of variance in systemizing was explained by SNPs. To our knowledge, there is no study examining heritability of the SQ-R in twins. One study, investigating the heritability of the reduced EQ (18 items) in 250 twin pairs identified a heritability of 0.32(10). The literature on the heritability of empathy and prosociality is inconsistent, with heritability estimates ranging from 0.20 (8) to 0.69 (39), although a meta-analysis of different studies identified a heritability estimate of 0.35 (95% CI – 0.21 – 0.41) (40). Our analysis therefore suggests that a third of the heritability can be attributed to common genetic variants. Like IQ (41), the heritability of empathy and prosociality behaviour changes with age (8). Our study did not find any significant

bioRxiv preprint doi: https://doi.org/10.1101/050682; this version posted February 23, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

differences in heritability between males and females. Further, the male-female genetic correlations for both the traits were high. However, the most significant SNPs ($P < 1x10^{-6}$) might have sex-specific contributions to the traits, since SNPs that are significant in one sex are not or are much less significant in the other sex ($P > 0.05$).

In conclusion, the current study provides the first narrow sense heritability for empathy and systemizing. While there is a very significant difference in scores on the two questionnaires between males and females, heritability is similar, with a high genetic correlation between the sexes. We also identified significant genetic correlations between empathy, systemizing and several psychiatric conditions, psychological traits, and measures of cognition. This global view of the genomic architecture of empathy and systemizing will allow us to better understand psychiatric conditions, but also to improve our knowledge of the biological bases of neurodiversity and brain evolution in humans.

## Methods

*Participants:* Participants were customers of 23andMe, Inc. a personal genetics company and are described in detail elsewhere(42, 43). 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 AAHRPP-accredited private institutional review board (http://www.eandireview.com). All participants completed the online version of the questionnaire on the 23andMe participant portal. The number of participants and participant overlap is provided in Table 1. Only participants who were primarily of European ancestry (97% European Ancestry) were selected for the analysis using existing methods(44). Unrelated individuals were selected using a segmental identity-by-descent algorithm(45).

**Table 1: Number of participants for each measure.**

| Measure | All | Female | Male |
|---|---|---|---|
| EQ | 46,861 | 24,543 | 22,318 |
| SQ-R | 51,564 | 25,501 | 26,063 |

bioRxiv preprint doi: https://doi.org/10.1101/050682; this version posted February 23, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

| | | | |
|---|---|---|---|
| SQ-R and EQ overlap | 38,889 | 20,127 | 18,762 |

*Measures:* Two online questionnaires were used in this study. The first, the Empathy Quotient (EQ) (1), is a self-report measure of empathy, and includes items relevant to both cognitive and affective empathy. It comprises 64 questions and has a good test-retest reliability (37). In this study, participants scored a maximum of 80 and a minimum of 0. The second measure is the Systemizing Quotient-Revised (SQ-R), which is self-report measure of systemizing drive, or interest in rule-based patterns (15). There are 75 items on the SQ-R, with a maximum score of 150 and a minimum score of 0. Scores on both the test are normally distributed.

*Genotyping, imputation and quality control:* DNA extraction and genotyping were performed on saliva samples by the National Genetic Institute, USA. Participants were genotyped using one of four different genotyping platforms. Participants were genotyped on one of four different platforms – V1, V2, V3 and V4. The V1 and V2 platforms have a total of 560,000 SNPs largely based on the Illumina HumanHap550+ BeadChip. The V3 platform has 950,000 SNPs based on the Illumina OmniExpress+ Beadchip and has custom content to improve the overlap with the V2 platform. The V4 platform is a fully customized array and has about 570,000 SNPs. All samples had a call rate greater than 98.5%. A total of 1,030,430 SNPs (including Insertion/Deletion or InDels) were genotyped across all platforms. Genotyped SNPs were filtered for quality control. Imputation was performed using the September 2013 release of the 1000 Genomes Project Phase 1reference haplotype phased using ShapeIt2. SNPs present only on platform V1, or in chromosome Y and mitochondrial chromosomes were excluded due to small sample sizes and unreliable genotype calling respectively. Next, using trio data, where available, SNPs that failed a parent offspring transmission test were excluded. SNPs were also excluded if they failed the Hardy-Weinberg Equilibrium Test at $P < 10^{-20}$, or had a genotype rate of less than 90%. Phasing was performed by 23andMe using an internal tool called Finch which implements the Beagle haplotype graph-based phasing algorithm. SNPs were excluded if they were not in Hardy-Weinberg equilibrium ($P < 10^{-20}$), had a call rate less than 95%, or had discrepancies in allele frequency compared to the reference European 1000 Genomes data (chi-squared $P < 10^{-15}$). Imputation was performed using Minimac25 using the September 2013 release of the 1000 Genomes Phase 1 reference haplotypes phased using Beagle4 (V3.3.1). For imputation, chromosomes were divided into segments of 10,000 genotyped SNPs, with overlaps of 200

bioRxiv preprint doi: https://doi.org/10.1101/050682; this version posted February 23, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

SNPs, and imputed against all-ethnicity 1000 Genomes haplotypes (excluding monomorphic and singleton sites) using Minimac25, using 5 rounds and 200 states for parameter estimation. We restricted the analyses to only SNPs that had a minor allele frequency (MAF) of at least 1%. After quality control, 9,955,952 SNPs were analysed. Genotyping, imputation, and preliminary quality control were performed by 23andMe.

*Genetic association:* We performed a linear regression assuming an additive model of genetic effects. Age and sex along with the first five ancestry principal components were included as covariates. Additionally, for each trait, we performed a male-only and a female-only linear regression analysis to identify sex-specific loci. Since we were performing three tests for each trait (male-only, female-only, and males and females combined with sex as a covariate), we used a threshold of $P < 1.66 \times 10^{-8}$ to identify significant SNPs for each trait. Leading SNPs in each loci were identified after pruning for LD ($r^2 > 0.8$) using SNAP (46). We calculated the variance explained by the top SNPs using the following formula:

$$R2g|c/(1 - R2c) = (t2/((n-k-1)+t2))x100$$

R2g|c/(1 − R2c) is the proportion of variance explained by the SNP after accounting for the effects of the covariates (four ancestry principal components, age, and, additionally, sex for the non-stratified analyses), t is the t-statistic of the regression co-efficient, k is the number of covariates, and n is the sample size.

*Genomic inflation factor, heritability, and functional enrichment:* We used Linkage Disequilibrium Score regression coefficient (LDSC) to calculate genomic inflation due to population stratification (25) (https://github.com/bulik/ldsc). The intercept for the SQ-R GWAS was 0.998 and the intercept for the EQ GWAS was 0.993 indicating that there was no unaccounted population stratification. Heritability and genetic correlation was performed using extended methods in LDSC (47). Difference in heritability between males and females was quantified using:

$$Z_{diff} = (h^2_{males} - h^2_{females}) / sqrt(SE^2_{males} + SE^2_{females})$$

where $Z_{diff}$ 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 $P < 0.05$. We identified enrichment in genomic functional elements for the traits using extended methods in LDSC (48).

bioRxiv preprint doi: https://doi.org/10.1101/050682; this version posted February 23, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

*Genetic correlations:* LDSC was also used to calculate genetic correlations. We restricted our analyses to only the non-stratified GWAS dataset due to the unavailability of sex-stratified GWAS data in the phenotypes investigated. We calculated initial genetic correlations using LD Hub (34) for schizophrenia (49), autism, ADHD, bipolar disorder, major depressive disorder, depressive symptoms, educational attainment (years of schooling.), NEO-Openness to experience, NEO-Conscientiousness, subjective wellbeing, and neuroticism. For anorexia nervosa (50), we used the latest PGC data freeze (https://www.med.unc.edu/pgc/results-and-downloads) which has more cases than the dataset available on LD Hub. In addition, we conducted also genetic correlation for extraversion (51) separately as the data was unavailable on LD Hub. For the anorexia and the extraversion analyses, the North West European LD scores were used and the intercepts were not constrained as the extent of participant overlap was unknown. We report significant lists if they have a Benjamini-Hochberg based FDR $q < 0.05$. FDR based correction was conducted for all the genetic correlations analyses combined (i.e. the genetic correlations for EQ and SQ-R combined) except for the quasi-replication analyses described next. In addition, we conducted quasi replication analyses of the systemizing – educational attainment genetic correlation using two additional measures of cognition: college completion and childhood cognition. These were conducted using LDHub and we did not include them in the FDR correction. For anorexia nervosa, we also conducted genetic correlation analyses using the sex-stratified EQ dataset as the participants in the anorexia nervosa dataset were predominantly females (68% females). A positive genetic correlation indicates that variants that contributes to either EQ or SQ-R increase the risk for conditions like schizophrenia and autism, whereas a negative genetic correlation indicates a decrease in risk.

*Gene-based analysis:* Gene based analyses for the non-stratified GWAS were performed using MetaXcan (52) using tissue weights derived from the GTEx project (53). MetaXcan uses summary GWAS data and tissue eQTL information derived from the GTEx project to impute gene expression levels for each gene. This allows for tissue specific prioritization of genes. For each trait, we ran gene-based analysis for ten neural tissues: anterior cingulate cortex (BA24), caudate basal ganglia, cerebellar hemisphere, cerebellum, cortex, frontal cortex (BA9), hippocampus, hypothalamus, nucleus accumbens basal ganglia, and putamen basal ganglia. We removed genes where there were 0 SNPs from our dataset, and genes that correlated poorly with predicted models of gene-expression ($R^2 < 0.01$) as implemented in the software. We report significant lists if they have an FDR < 0.05. We confirmed that LD,

bioRxiv preprint doi: https://doi.org/10.1101/050682; this version posted February 23, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

SNP density, SNP number, and gene size do not confound gene identification (Supplementary Figures 20 and 21).

*Data Availability:* Summary GWAS data for the EQ and the SQ-R is available for qualified researchers from 23andMe, Inc. Top SNPs can be visualized here: https://ghfc.pasteur.fr/eacf331f0ffc35d4b482f1d15a887d3b

**Acknowledgements**

We thank Richard Bethlehem, Florina Uzefovsky, Paula Smith, and Carrie Allison for discussions of the results. We are grateful to Brendan Bulik-Sullivan, Hillary Finucane, and Donna Werling for their help with the analytical methods. This study was funded by grants from the Medical Research Council, the Wellcome Trust, the Autism Research Trust, the Institut Pasteur, the CNRS and the University Paris Diderot. VW is funded by St. John's College, Cambridge, and Cambridge Commonwealth Trust. The research was funded and supported by 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. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. We would like to thank the research participants and employees of 23andMe for making this work possible. This work was supported by the National Human Genome Research Institute of the National Institutes of Health (grant number R44HG006981).

**Disclosure**

NL and DAH are employed by 23andMe and own stock options in the company. VW, RT, BC, TB, and SBC have no biomedical or interests or conflicts of interest to disclose.

**References**

1. Baron-Cohen S, Wheelwright S (2004): The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *J Autism Dev Disord*. 34: 163–75.

2. Derntl B, Seidel E-M, Schneider F, Habel U (2012): How specific are emotional deficits?

bioRxiv preprint doi: https://doi.org/10.1101/050682; this version posted February 23, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A comparison of empathic abilities in schizophrenia, bipolar and depressed patients. *Schizophr Res*. 142: 58–64.

3. Bora E, Gökçen S, Veznedaroglu B (2008): Empathic abilities in people with schizophrenia. *Psychiatry Res*. 160: 23–9.

4. Michaels TM, Horan WP, Ginger EJ, Martinovich Z, Pinkham AE, Smith MJ (2014): Cognitive empathy contributes to poor social functioning in schizophrenia: Evidence from a new self-report measure of cognitive and affective empathy. *Psychiatry Res*. 220: 803–10.

5. Lehmann A, Bahçesular K, Brockmann E-M, Biederbick S-E, Dziobek I, Gallinat J, Montag C (2014): Subjective experience of emotions and emotional empathy in paranoid schizophrenia. *Psychiatry Res*. 220: 825–33.

6. Weightman MJ, Air TM, Baune BT (2014): A Review of the Role of Social Cognition in Major Depressive Disorder. *Front Psychiatry*. 5: 179.

7. Thoma P, Schmidt T, Juckel G, Norra C, Suchan B (2015): Nice or effective? Social problem solving strategies in patients with major depressive disorder. *Psychiatry Res*. . Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/26051176.

8. Davis MH, Luce C, Kraus SJ (1994): The heritability of characteristics associated with dispositional empathy. *J Pers*. 62: 369–91.

9. Emde RN, Plomin R, Robinson JA, Corley R, DeFries J, Fulker DW, *et al.* (1992): Temperament, emotion, and cognition at fourteen months: the MacArthur Longitudinal Twin Study. *Child Dev*. 63: 1437–55.

10. Hatemi PK, Smith K, Alford JR, Martin NG, Hibbing JR (2015): The genetic and environmental foundations of political, psychological, social, and economic behaviors: a panel study of twins and families. *Twin Res Hum Genet*. 18: 243–55.

11. Warrier V, Baron-Cohen S, Chakrabarti B (2013): Genetic variation in GABRB3 is associated with Asperger syndrome and multiple endophenotypes relevant to autism. *Mol Autism*. 4: 48.

12. Uzefovsky F, Shalev I, Israel S, Edelman S, Raz Y, Perach-Barzilay N, *et al.* (2014): The dopamine D4 receptor gene shows a gender-sensitive association with cognitive empathy: evidence from two independent samples. *Emotion*. 14: 712–21.

bioRxiv preprint doi: https://doi.org/10.1101/050682; this version posted February 23, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

13. Chakrabarti B, Dudbridge F, Kent L, Wheelwright S, Hill-Cawthorne G, Allison C, *et al.* (2009): Genes related to sex steroids, neural growth, and social-emotional behavior are associated with autistic traits, empathy, and Asperger syndrome. *Autism Res*. 2: 157–77.

14. Baron-Cohen S, Richler J, Bisarya D, Gurunathan N, Wheelwright S (2003): 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*. 358: 361–74.

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

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

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

18. de Moor MHM, van den Berg SM, Verweij KJH, Krueger RF, Luciano M, Arias Vasquez A, *et al*. (2015): Meta-analysis of Genome-wide Association Studies for Neuroticism, and the Polygenic Association With Major Depressive Disorder. *JAMA psychiatry*. 72: 642–50.

19. Power RA, Steinberg S, Bjornsdottir G, Rietveld CA, Abdellaoui A, Nivard MM, *et al.* (2015): Polygenic risk scores for schizophrenia and bipolar disorder predict creativity. *Nat Neurosci*. 18: 953–955.

20. Clarke T-K, Lupton MK, Fernandez-Pujals AM, Starr J, Davies G, Cox S, *et al*. (2015): Common polygenic risk for autism spectrum disorder (ASD) is associated with cognitive ability in the general population. *Mol Psychiatry*. 21: 419–25.

21. Baron-Cohen S (2010): Empathizing, systemizing, and the extreme male brain theory of autism. *Prog Brain Res*. 186: 167–75.

22. Billington J, Baron-Cohen S, Wheelwright S (2007): Cognitive style predicts entry into physical sciences and humanities: Questionnaire and performance tests of empathy and

bioRxiv preprint doi: https://doi.org/10.1101/050682; this version posted February 23, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

systemizing. *Learn Individ Differ*. 17: 260–268.

23. Chapman E, Baron-Cohen S, Auyeung B, Knickmeyer R, Taylor K, Hackett G (2006): Fetal testosterone and empathy: evidence from the empathy quotient (EQ) and the "reading the mind in the eyes" test. *Soc Neurosci*. 1: 135–48.

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

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

26. Auyeung B, Lombardo M V, Baron-Cohen S (2013): Prenatal and postnatal hormone effects on the human brain and cognition. *Pflugers Arch*. 465: 557–71.

27. Horan WP, Reise SP, Kern RS, Lee J, Penn DL, Green MFStructure and correlates of self-reported empathy in schizophrenia. *J Psychiatr Res*. 66-67: 60–6.

28. Warrier V, Grasby K, Uzefovsky F, Toro R, Smith P, Chakrabarti B, *et al.* (2016): Genome-wide meta-analysis of cognitive empathy: heritability, and correlates with sex, neuropsychiatric conditions and brain anatomy. *bioRxiv*. .

29. Larson F V, Lai M-C, Wagner AP, MRC AIMS Consortium, Baron-Cohen S, Holland AJ (2015): Testing the "Extreme Female Brain" Theory of Psychosis in Adults with Autism Spectrum Disorder with or without Co-Morbid Psychosis. *PLoS One*. 10: e0128102.

30. Brosnan M, Ashwin C, Gamble T (2013): Greater Empathizing and reduced Systemizing in people who show a jumping to conclusions bias in the general population: Implications for psychosis. *Psychosis*. 5: 71–81.

31. Trampush JW, Yang MLZ, Yu J, Knowles E, Davies G, Liewald DC, *et al.* (2017): GWAS meta-analysis reveals novel loci and genetic correlates for general cognitive function: a report from the COGENT consortium. *Mol Psychiatry*. . Retrieved February 14, 2017, from http://www.nature.com/doifinder/10.1038/mp.2016.244.

32. Melchers MC, Li M, Haas BW, Reuter M, Bischoff L, Montag C (2016): Similar Personality Patterns Are Associated with Empathy in Four Different Countries. *Front Psychol*. 7: 290.

bioRxiv preprint doi: https://doi.org/10.1101/050682; this version posted February 23, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

33. Bos EH, Snippe E, de Jonge P, Jeronimus BF, Keyes C, Perkins R, *et al*. (2016): Preserving Subjective Wellbeing in the Face of Psychopathology: Buffering Effects of Personal Strengths and Resources. (A. Kavushansky, editor) *PLoS One*. 11: e0150867.

34. Zheng J, Erzurumluoglu AM, Elsworth BL, Kemp JP, Howe L, Haycock PC, *et al.* (2016): LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics*. 33: 272–279.

35. Ashton MC, Lee K, Vernon PA, Jang KL (2000): Fluid Intelligence, Crystallized Intelligence, and the Openness/Intellect Factor. *J Res Pers*. 34: 198–207.

36. Moutafi J, Furnham A, Crump J (2006): What facets of openness and conscientiousness predict fluid intelligence score? *Learn Individ Differ*. 16: 31–42.

37. Lawrence EJ, Shaw P, Baker D, Baron-Cohen S, David AS (2004): Measuring empathy: reliability and validity of the Empathy Quotient. *Psychol Med*. 34: 911–9.

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

39. Knafo-Noam A, Uzefovsky F, Israel S, Davidov M, Zahn-Waxler C (2015): The prosocial personality and its facets: genetic and environmental architecture of mother-reported behavior of 7-year-old twins. *Front Psychol*. 6: 112.

40. Knafo A, Uzefovsky F (2013): Variation in empathy: The interplay of genetic and environmental factors. In: Legerstee M, Haley DW, Bornstein MH, editors. *infant mind Orig Soc Brain*. The Guilford Press, pp 97 – 121.

41. Bouchard TJ (2013): The Wilson Effect: the increase in heritability of IQ with age. *Twin Res Hum Genet*. 16: 923–30.

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

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

bioRxiv preprint doi: https://doi.org/10.1101/050682; this version posted February 23, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

44. Eriksson N, Tung JY, Kiefer AK, Hinds DA, Francke U, Mountain JL, Do CB (2012): Novel associations for hypothyroidism include known autoimmune risk loci. *PLoS One*. 7: e34442.

45. Henn BM, Hon L, Macpherson JM, Eriksson N, Saxonov S, Pe'er I, Mountain JL (2012): Cryptic distant relatives are common in both isolated and cosmopolitan genetic samples. *PLoS One*. 7: e34267.

46. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PIW (2008): SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics*. 24: 2938–9.

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

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

49. Ripke S, Neale BM, Corvin A, Walters JTR, Farh K-H, Holmans PA, *et al.* (2014): Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 511: 421–7.

50. Duncan L, Yilmaz Z, Walters R, Goldstein J, Anttila V, Bulik-Sullivan B, *et al.* (2016): Genome-Wide Association Study Reveals First Locus for Anorexia Nervosa and Metabolic Correlations. *bioRxiv*. .

51. van den Berg SM, de Moor MHM, Verweij KJH, Krueger RF, Luciano M, Arias Vasquez A, *et al.* (2016): Meta-analysis of Genome-Wide Association Studies for Extraversion: Findings from the Genetics of Personality Consortium. *Behav Genet*. 46: 170–82.

52. Barbeira A, Shah KP, Torres JM, Wheeler HE, Torstenson ES, Edwards T, *et al.* (2016): *MetaXcan: Summary Statistics Based Gene-Level Association Method Infers Accurate PrediXcan Results*. *bioRxiv*. . Retrieved April 12, 2016, from http://biorxiv.org/content/early/2016/03/23/045260.abstract.

53. Ardlie KG, Deluca DS, Segre A V., Sullivan TJ, Young TR, Gelfand ET, *et al.* (2015): The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science (80- )*. 348: 648–660.

bioRxiv preprint doi: https://doi.org/10.1101/050682; this version posted February 23, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

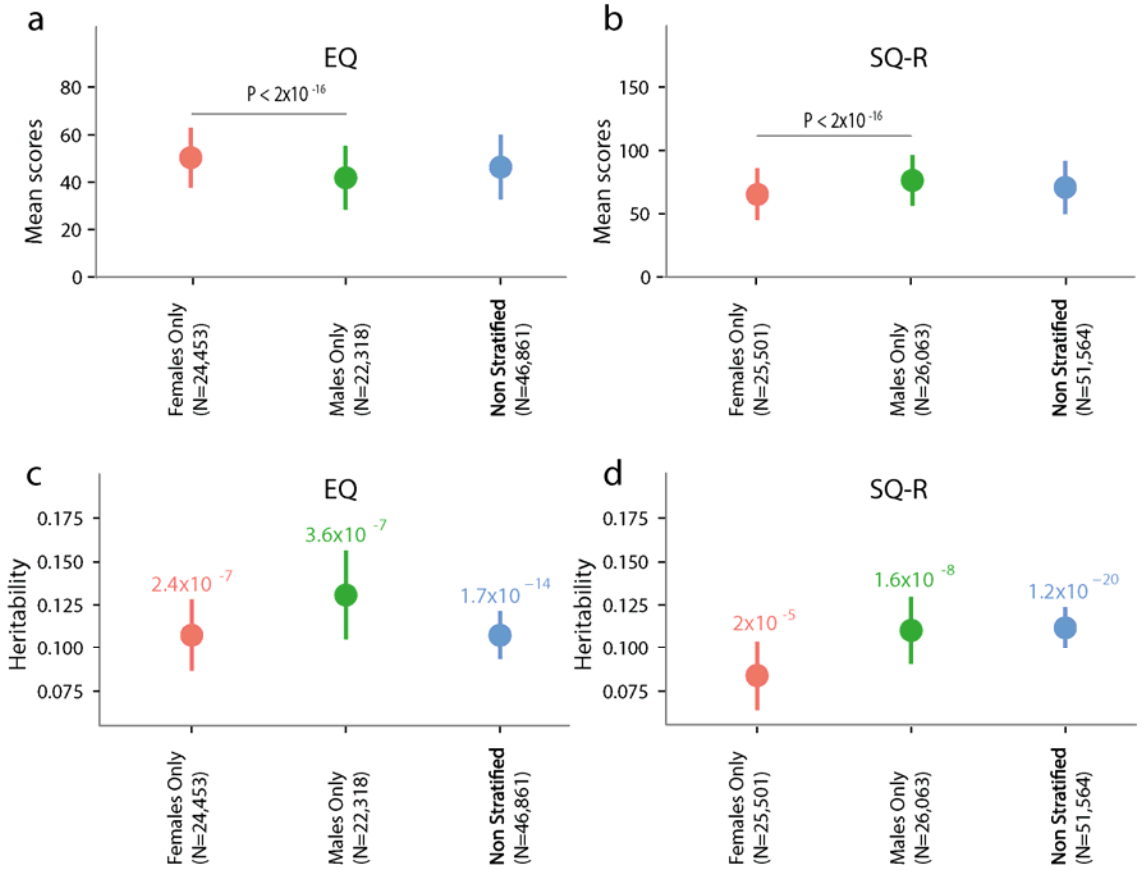


**Figure 1: Mean scores and heritability estimates for SQ-R and EQ**. *Mean scores and standard deviations for scores on the EQ (a) and the SQ-R (b). The effect sizes of the difference between males and females on the EQ scores and the SQ-R scores were Cohen's d= 0.65 and Cohen's d = 0.54, respectively. Mean estimates and standard errors for*

bioRxiv preprint doi: https://doi.org/10.1101/050682; this version posted February 23, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

*heritability for scores on the EQ (c) and the SQ-R (d). Numbers on top of the graphs represent P-values for each heritability estimate.*

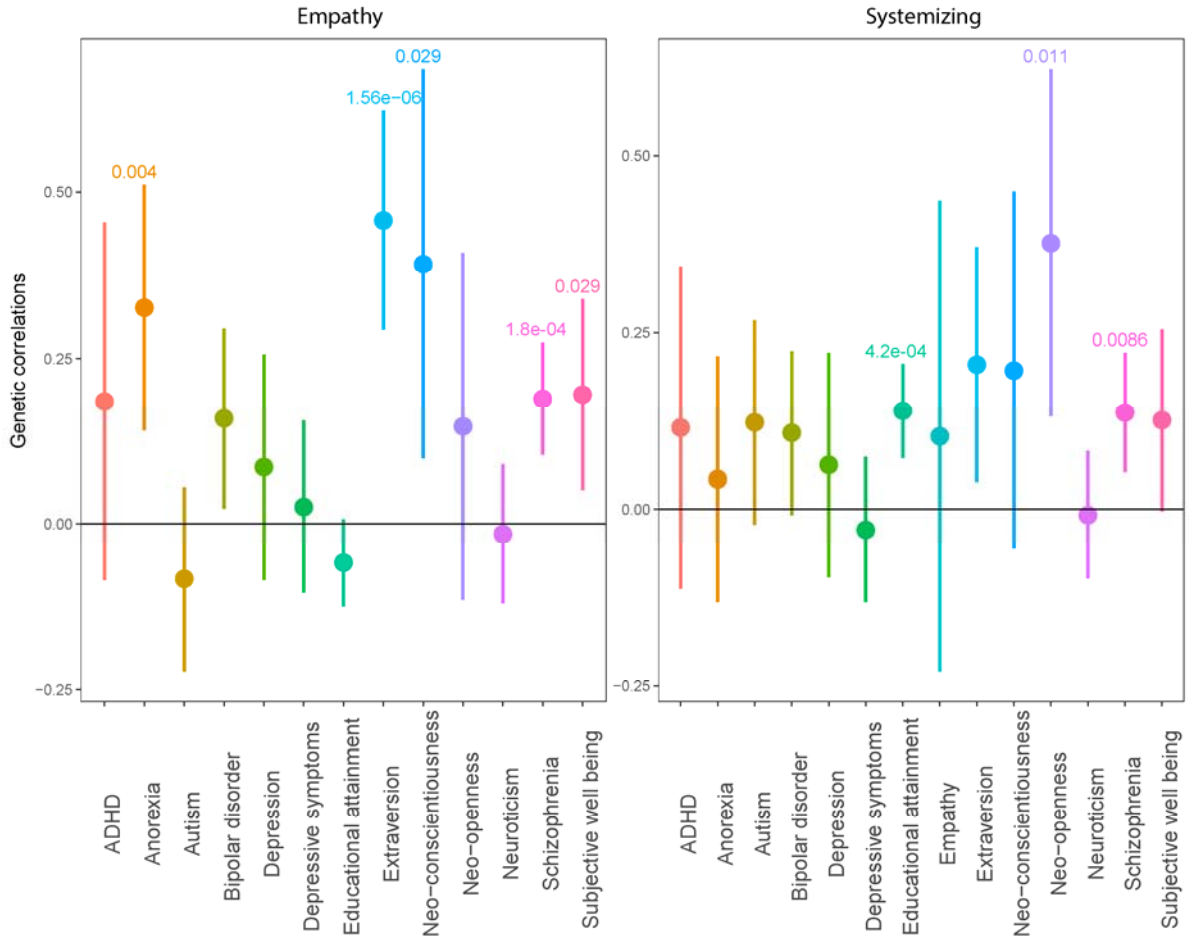


**Figure 2: Genetic correlations between SQ-R and EQ and other conditions**. *Mean and 95% confidence intervals shown for genetic correlations between empathy, systemizing and other conditions. FDRadjusted P-values are provided for significant correlations. The sample sizes for the traits and references are provided in Supplementary Table 7. ADHD – Attention Deficit Hyperactivity Disorder, Depression – Major Depressive Disorder, Educational attainment – Years of Schooling (Okbay et al., 2016).*