bioRxiv preprint doi: https://doi.org/10.1101/050682; this version posted November 14, 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 The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.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 self-reported empathy: correlations with autism, schizophrenia, and anorexia nervosa

Short title: GWAS of self-reported empathy

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.

bioRxiv preprint doi: https://doi.org/10.1101/050682; this version posted November 14, 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 The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.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

Empathy is the ability to recognize and respond to the emotional states of other individuals. It is an important psychological process that facilitates navigating social interactions and maintaining relationships, which are important for wellbeing. Several psychological studies have identified difficulties in both self-report and performance-based measures of empathy in a range of psychiatric conditions. To date, no study has systematically investigated the genetic architecture of empathy using genome-wide association studies (GWAS). Here we report the results of the largest GWAS of empathy to date using a well-validated self-report measure of empathy, the Empathy Quotient (EQ), in 46,861 research participants from 23andMe, Inc. We identify 11 suggestive loci ($P < 1x10^{-6}$), though none were significant at $P < 2.5x10^{-8}$ after correcting for multiple testing. The most significant SNP was identified in the non-stratified analysis (rs4882760; $P = 4.29x10^{-8}$), and is an intronic SNP in *TMEM132C*. The EQ had a modest but significant narrow-sense heritability (0.11 ± 0.014; $P = 1.7x10^{-14}$). As predicted, based on earlier work, we confirmed a significant female-advantage on the EQ ($P < 2x10^{-16}$ Cohen's d = 0.65). We identified similar SNP heritability and high genetic correlation between the sexes. Also, as predicted, we identified a significant negative genetic correlation between autism and the EQ ($r_g = -0.27 \pm 0.07$, $P = 1.63x10^{-4}$). We also identified a significant positive genetic correlation between the EQ and risk for schizophrenia ($r_g = 0.19 \pm 0.04$; $P= 1.36x10^{-5}$), risk for anorexia nervosa ($r_g = 0.32 \pm 0.09$; $P = 6x10^{-4}$), and extraversion ($r_g = 0.45 \pm 0.08$; $5.7x10^{-8}$). This is the first GWAS of self-reported empathy. The results suggest that the genetic variations associated with empathy also play a role in psychiatric conditions and psychological traits.

Keywords: Empathy, GWAS, schizophrenia, autism, anorexia nervosa

bioRxiv preprint doi: https://doi.org/10.1101/050682; this version posted November 14, 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.

Introduction

Empathy is the ability to identify other people's thoughts, intentions, desires, and feelings, and to respond to others' mental states with an appropriate emotion[1]. It plays an important role in social interaction by facilitating both making sense of other people's behaviour and in responding appropriately to their behaviour. For these reasons, it is considered a key component of prosocial behaviour, social cooperation, and social cognition[2]. Aspects of empathy are observed in humans and other animals and is thought to have evolved to support a range of prosocial behaviour and cooperative behaviour[2].

Differences in various fractions empathy have been observed in several psychiatric conditions including autism[1], bipolar disorder[3], schizophrenia[4–6], and major depressive disorder[3,7,8]. Two major fractions of empathy include affective empathy (the drive to respond to another's mental state with an appropriate emotion) and cognitive empathy (the ability to recognize another's mental state). These differences vary between psychiatric conditions: for example, individuals with schizophrenia are more likely to report higher personal distress and emotional contagion[6], whereas individuals with autism are likely to show difficulties with cognitive empathy but not affective empathy[1,9]. These may reflect a causal risk mechanisms where alterations in empathy contribute to higher risk for developing a psychiatric condition. Equally, differences in empathy may also be due to the presence of a psychiatric condition, which may not allow individuals to understand and respond to another person's mental state effectively.

Whilst empathy is clearly shaped by early experience, parenting, and other social factors, different lines of evidence suggest that empathy is partly biological. Empathy is modestly heritable (approximately a third of the variance is heritable)[10–12], and a few candidate gene association studies have investigated the role of various genes in empathy[13–15]. In addition, several studies have identified a role for the oxytocinergic and the foetal testosterone systems in modulating empathy[2,16–18]. Neuroimaging studies have identified distinct brain regions implicated in different aspects of empathy including the amygdala and the ventromedial prefrontal cortex[19,20]. Empathy also shows a marked sex difference: females, on average, score higher on different measures of empathy[1,21]. A longitudinal study suggests that this female advantage grows larger with age[22]. Sex differences in the mind arise from a combination of innate biological differences, cultural, and environmental differences. Studies in infant humans have identified sex differences in the developmental precursors to empathy, such as neonatal preferences to faces over objects[23] when environmental and

bioRxiv preprint doi: https://doi.org/10.1101/050682; this version posted November 14, 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.

cultural influences are minimal, lending support to the idea that sex differences in empathy are at least partly biological[24].

Because empathy difficulties are found in a range of psychiatric conditions, it is an important phenotype for investigation. Understanding the biological networks that partly determine empathy may help us understand how it contributes to psychiatric phenotypes, an approach that has been used for other traits such as neuroticism[25], creativity[26], and cognitive ability[27]. We investigate the genetic architecture of self-reported empathy using the Empathy Quotient (EQ)[1]. The EQ is listed in the Research Domain Criteria (RDoC)[28] as a self-report measure under the domain of 'Understanding Mental States' (https://www.nimh.nih.gov/research-priorities/rdoc/units/self-reports/151133.shtml). The EQ is widely used, has excellent test-retest reliability ($r = 0.83$, $P = 0.0001$)[29], high internal consistency (Cronbach's alpha ~ 0.9)[30,31] and is significantly correlated with factors in the Interpersonal Reactivity Index and the Toronto Empathy Questionnaire, two other measures of empathy, suggesting good concurrent validity[29,32]. Psychometric analysis of the EQ in 3,334 individuals from the general population suggests that the EQ is a good measure of empathy which can be measured across a single dimension[33]. By focusing on a self-report measure of empathy, we were able to obtain phenotypic and genetic data from a large number of participants, increasing the statistical power of the study. Previous work from our lab investigated the genetic correlates of cognitive empathy using a specific performance measure – the 'Reading the Mind in the Eyes' Test (the Eyes Test)[34]. The Eyes Test has a low correlation with the EQ ($r \sim 0.10$)[30,35] and measures only one facet of empathy, namely cognitive empathy. Cognitive empathy is also referred to as employing a 'theory of mind', or 'mentalizing'. The EQ includes items that measure cognitive empathy, but others that measure affective empathy. Yet other items on the EQ involve both cognitive and affective empathy.

In this study, we aim to answer three questions: 1. What is the genetic architecture of empathy? 2. Is empathy genetically correlated to various psychiatric conditions, psychological traits, and education? and 3. Is there a genetic contribution to sex differences in empathy? We performed sex-stratified and non-stratified genome-wide association analyses of empathy in research participants from 23andMe, a personalized genetics company. We calculated the narrow sense heritability explained by all the SNPs tested, and investigated sex differences. Finally, we conducted genetic correlation analyses with six psychiatric

bioRxiv preprint doi: https://doi.org/10.1101/050682; this version posted November 14, 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.

conditions (anorexia, ADHD, autism, bipolar disorder, major depressive disorder, and schizophrenia), psychological traits, and educational attainment.

Methods

Participants: Research Participants were drawn from the customer base of 23andMe, Inc. a personal genetics company and are described in detail elsewhere[36,37]. There were 46,861 participants (24,543 females and 22,318 males). All participants included in the analyses 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 accessible via the research tab of their password protected 23andMe personal online account. Only participants who were primarily of European ancestry (97% European Ancestry) were selected for the analysis using existing methods[38]. Unrelated individuals were selected using a segmental identity-by-descent algorithm[39].

Measures: The Empathy Quotient (EQ)[1], is a self-report measure of empathy, and includes items relevant to both cognitive and affective empathy. It comprises 60 questions and has a good test-retest reliability[29]. 20 questions are filler questions, of the remaining 40 questions participants can score a maximum of 2 points and a minimum of 0 point per question. Therefore, in this study, participants scored a maximum of 80 and a minimum of 0.

Genotyping, imputation and quality control: DNA extraction and genotyping were performed on saliva samples by the National Genetic Institute, USA. 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. Imputation was performed using the March 2012 (v3) release of the 1000 Genomes Phase 1 reference haplotypes. First, we used Beagle (version 3.3.1)[40] to phase batches of 8000-9000 individuals across chromosomal segments of no more than 10,000 genotyped SNPs, with overlaps of 200 SNPs. SNPs were excluded if they were not in Hardy-Weinberg equilibrium

bioRxiv preprint doi: https://doi.org/10.1101/050682; this version posted November 14, 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.

($P < 10^{-20}$), had a genotype call rate less than 95%, or had discrepancies in allele frequency compared to the reference European 1000 Genomes data (chi-squared $P < 10^{-15}$). We then imputed each phased segment against all-ethnicity 1000 Genomes haplotypes (excluding monomorphic and singleton sites) using Minimac2[41], 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%. For genotyped SNPs, those 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 from all research participants in the 23andMe dataset, where available, SNPs that failed a parent offspring transmission test were excluded. For imputed SNPs, we excluded SNPs with average $r^2 < 0.5$ or minimum $r^2 < 0.3$ in any imputation batch, as well as SNPs that had strong evidence of an imputation batch effect. The batch effect test is an F test from an ANOVA of the SNP dosages against a factor representing imputation batch; we excluded results with $P < 10^{-50}$. 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, we performed a male-only and a female-only linear regression analysis to identify sex-specific loci. Since we were performing two independent tests for each trait (male-only and female-only, and males and females combined with sex as a covariate which is equivalent to a meta-analysis of the two sex-stratified GWAS), we used a threshold of $P < 2.5\text{x}10^{-8}$ ($5\text{x}10^{-8}/2$) to identify significant SNPs. Leading SNPs in each locus was identified after pruning for LD ($r^2 > 0.8$) using Plink version 1.9. We calculated the variance explained by the top SNPs using a previously used formula[42]:

$$\frac{R^2_{g|c}}{1 - R^2_c} = \frac{t^2}{n - k - 1} \times 100$$

$\frac{R^2_{g|c}}{1-R^2_c}$ 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. Winner's curse correction was conducted using FDR Inverse Quantile Transformation[43].

bioRxiv preprint doi: https://doi.org/10.1101/050682; this version posted November 14, 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.

Genomic inflation factor, heritability, and functional enrichment: We used Linkage Disequilibrium Score regression coefficient (LDSR) to calculate genomic inflation due to population stratification[44] (https://github.com/bulik/ldsc). Heritability and genetic correlation was performed using extended methods in LDSR[45]. Difference in heritability between males and females was quantified using[46]:

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

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 $P < 0.05$. This We identified enrichment in genomic functional elements for the traits by partitioning heritability performed in LDSR[47]. In addition to the baseline partitions we conducted four additional enrichment analyses. Enrichment for CNS specific histone marks was conducted using cell type specific partitioned heritability analysis. For genes that are intolerant to loss-of-function mutations, we identified gene boundaries of genes with probability of loss-of-function intolerance scores > 0.9 from the Exome Aggregation Consortium[48], and conducted partitioned heritability analysis for all common SNPs within the gene boundaries identified. Similarly, for sex-differentially enriched genes, we identified gene boundaries of genes with sex differential expression in cerebral cortex and associates structures (Brain Other)[49] **(Supplementary Table 8)**. We divided this into two separate lists – genes with higher expression in males and genes with higher expression in females with an FDR corrected P-value < 0.05. Partitioned heritability analyses were conducted to identify enrichment using LDSR.

Genetic correlations: LDSR 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[50] for schizophrenia[51], 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[52], and

bioRxiv preprint doi: https://doi.org/10.1101/050682; this version posted November 14, 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.

ADHD[53], we used the data available from the PGC webpage (https://www.med.unc.edu/pgc/results-and-downloads) to conduct genetic correlation analyses as these are in larger samples and, consequently, have greater statistical power than the datasets available on LD Hub. For autism, we used summary statistics from the PGC-iPSYCH meta-analysis[54], details of which are provided in the **Supplementary Note**. We also conducted genetic correlation for extraversion[55] separately as the data was unavailable on LD Hub. For the anorexia nervosa, autism 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 Bonferroni corrected $P < 0.05$, which we acknowledge is conservative. For anorexia nervosa and autism, we also conducted genetic correlation analyses using the sex-stratified EQ dataset due to the significant sex-differences observed in these conditions. We correct for these using Bonferroni correction, and report significant correlations at $P < 0.05$.

Gene-based analysis: Gene based analyses for the non-stratified GWAS were performed using MAGMA[56], which integrates LD information between SNPs to prioritize genes. Genes were significant if they had a Bonferroni corrected $P < 0.05$. In addition, we also investigated enrichment in Gene Ontology terms using MAGMA.

Genome-wide colocalization: Pairwise genome-wide colocalization analyses were conducted using GWAS-PW[57] by dividing the genomes into segments containing approximately 5,000 SNPs each. We considered the posterior probability of model 3 i.e. the model wherein SNPs in the same locus influence both the traits. We used a rigorous threshold of posterior probability > 0.95 to identify significant loci that influenced both the traits. We conducted pairwise colocalization for empathy (non-stratified), and schizophrenia[51], anorexia nervosa[52], and autism.

Data Availability: Summary statistics for the EQ GWAS can be requested directly from 23andMe, and will be made available to qualified researchers subject to the terms of a data transfer agreement with 23andMe that protects the privacy of the 23andMe research participants. Please contact David Hinds (dhinds@23andMe.com) for more information. Top SNPs can be visualized here: https://ghfc.pasteur.fr/eq/.

bioRxiv preprint doi: https://doi.org/10.1101/050682; this version posted November 14, 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.

Results

Phenotype description

To understand the genetic architecture of empathy, we collaborated with 23andMe to conduct a Genome Wide Association Study (GWAS) of empathy (n = 46,861) using the EQ, which was normally distributed. A flow chart of the study protocol is shown in **Figure 1**. The mean score for all participants was 46.4 (sd = 13.7) on a total of 80 on the EQ, which is similar to the mean score reported in 90 typical participants in the first study describing the EQ (42.1, sd =10.6)[1]. The mean age of the participants was 48.9 (sd = 15.7). Females scored higher than males on the EQ (41.9±13.5 in males, 50.4±12.6 in females) **(Figure 2)**, as previously observed[1]. There was significant age effect, with scores increasing with age (Beta = 0.08±0.003; P = 3.3×10^{-104}) and a significant sex effect with females scoring higher than males (Beta = 8.4±0.11; P ~ 0).

Insert Figure 1 here

Genome-wide association analyses

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 **(Methods)**. We corrected for the three different tests using a conservative threshold of P = 2.5×10^{-8}. LD score regression coefficient suggested non-significant genomic inflation due to population stratification **(Supplementary Figure 1 – 3)**. We did not identify any genome wide significant SNPs **(Supplementary Figures 1 - 3 and Supplementary Table 1)**. We identified 11 suggestive loci ($P < 1 \times 10^{-6}$) in the three GWAS. The most significant SNP was identified in the non-stratified analysis (rs4882760; P = 4.29×10^{-8}), and is an intronic SNP in *TMEM132C*. Regional association plots for all suggestive loci are provided in **Supplementary Figure 4**.

To investigate if the top SNPs from the EQ also contribute to cognitive empathy as measured by the Eyes Test, we conducted SNP lookup of all the 11 suggestive loci in the Eyes Test GWAS. None of the 11 loci were significant in the Eyes Test GWAS, and only 7 out of the 11 SNPs had concordant effect directions in the two traits (P = 0.54; two-sided binomial sign test).

bioRxiv preprint doi: https://doi.org/10.1101/050682; this version posted November 14, 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.

The most significant SNP in each GWAS analysis explained 0.06 – 0.13% of the total variance **(Supplementary Table 2).** However, this reduced to 0.0006 – 0.016% after correcting for winner's curse **(Supplementary Table 2)**.

Gene-based association, heritability, and enrichment in functional categories

Gene based analysis identified two significant genes for the EQ: *SEMA6D* ($P = 9.14 \times 10^{-7}$), and *FBN2* ($P = 1.68 \times 10^{-6}$) **(Supplementary Table 3)**. Analysis for enrichment in Gene Ontology (GO) terms did not identify any significant enrichment **(Supplementary Table 4)**. The most significant GO process was negative regulation of neurotransmitter secretion.

We used LDSR[44] to calculate the heritability explained by all the SNPs tested (Methods) and identified a heritability of 0.11 ± 0.014 for the EQ ($P = 1.7 \times 10^{-14}$) **(Figure 2, Supplementary Table 5)**. Partitioning heritability by functional categories did not identify any significant enrichment after correcting for multiple testing **(Supplementary Table 6)**. We also investigated if there was an enrichment in heritability for histone marks in cells in the CNS[47], but did not find a significant enrichment (enrichment = 3.67 ± 1.45; $P = 0.077$).

Recent studies have identified an enrichment of associations in or near genes that are extremely intolerant to loss-of-function variation in schizophrenia[58], autism[59,60], and developmental disorders[61], conditions that are often accompanied by difficulties in social behaviour and empathy. We investigated if there was a significant enrichment in GWAS signal for the EQ in these genes that are extremely intolerant to loss-of-function variation. We did not identify a significant enrichment after correction for multiple testing (proportion $h^2_{SNP} = 0.19$, proportion SNP = 0.09, enrichment = 1.83 ± 0.42; $P = 0.044$).

insert Figure 2 here

Sex differences

Sex differences in empathy[62] may reflect genetic as well as non-genetic factors (such as prenatal steroid hormones, and postnatal learning)[63]. In our dataset, there was a significant female advantage on the EQ ($P < 2 \times 10^{-16}$ Cohen's d = 0.65) **(Figure 2)**. To investigate the biological basis for the sex-difference observed in the traits, we first tested the heritability of the sex-stratified GWAS analyses for the EQ. Our analyses revealed no significance

bioRxiv preprint doi: https://doi.org/10.1101/050682; this version posted November 14, 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.

difference between the heritability in the males-only and the females-only datasets (P = 0.48 for male-female difference in the EQ) **(Figure 2 and Supplementary Table 5)**. Additionally, there was a high genetic-correlation between the males-only and females-only GWAS (r_g = 0.82±0.16, P = 2.34×10^{-7}), indicating a high degree of similarity in the genetic architecture of the traits in males and females. This was not significantly different from 1 (P = 0.13, one-sided Wald Test). We investigated the heterogeneity in the 11 SNPs of suggestive significance in both the sexes using Cochran's Q-Test, and did not identify significant heterogeneity **(Supplementary Table 7)**.

Sex differences may also arise by differential expression of specific genes in different neural tissues at different developmental stages[49,64]. This could be due to multiple factors, including sex-specific transcription factors and sex-specific DNA methylation. We investigated this by performing enrichment analysis of the non-stratified GWAS in genes with higher expression in either males or females in cortical tissue samples **(Supplementary Table 8)**. We did not identify a significant enrichment for either genes with higher expression in males (enrichment = 1.95±0.70, P = 0.17) or females (enrichment = 0.28±0.84, P = 0.39).

Genetic correlations

To investigate how the EQ correlates with psychiatric conditions, psychological traits and educational attainment, we performed genetic correlation **(Methods)** with six 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 9)**. With psychiatric conditions, three genetic correlations were significant following Bonferroni correction: EQ-autism (r_g = -0.27 ± 0.07, P = 1.63×10^{-4}), EQ-schizophrenia (r_g = 0.19±0.04; P = 1.36×10^{-5}) and EQ-anorexia nervosa (r_g = 0.32±0.09; P = 6×10^{-4}) (**Figure 3**).

As anorexia nervosa is primarily diagnosed in women, and autism is primarily diagnosed in men, we further tested sex-specific correlations. After Bonferroni correction, we identified significant genetic correlations between the EQ in females (EQ-F) and anorexia (r_g = 0.48±0.12; P = 8.46×10^{-5}) and the EQ in males (EQ-M) and autism (r_g = -0.3±0.08, P = 3×10^{-4}).

bioRxiv preprint doi: https://doi.org/10.1101/050682; this version posted November 14, 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.

With psychological traits, we identified one significant correlation after Bonferroni correction: EQ with extraversion (r_g = 0.45±0.08; P = 5.76x10^{-8}). Additionally, we identified two nominally significant correlations: the EQ with subjective wellbeing (r_g = 0.19±0.07; P = 7.8x10^{-3}), and NEO-conscientiousness (r_g = 0.39±0.14; P = 8.8x10^{-03}) (**Figure 3**). All three correlations were in the predicted direction as studies have identified a positive phenotypic correlation between all three traits and the EQ[31,65]. We previously reported a small positive correlation between the EQ and the Eyes Test (r_g = 0.18±0.06; P = 0.007)[34], mirroring previous reported estimates of phenotypic correlation in the general population[35] and estimates in our database from 916 neurotypical adults (r = 0.11±0.032; P = 0.003, Pearson correlation).

Bayesian genomic colocalization

As there were significant genetic correlations between the EQ, anorexia nervosa schizophrenia, and autism, we investigated if there are genomic regions that influence both empathy and one of the psychiatric conditions (colocalization) by estimating the Bayesian posterior probability. We did not identify any regions associated with empathy and the three conditions. The most significant region identified in this analysis was in Chr11p12, posterior probability = 0.78 **(Supplementary Figure 5)** in the empathy-anorexia analysis. The most significant SNPs in this region for both anorexia and empathy are intronic SNPs in the gene *LRRC4C*, which is implicated in excitatory synapse development[66,67]. Further, this gene is highly intolerant to loss-of-function mutations (probability of Loss-of-function Intolerance = 0.95). We did not identify any eQTLs in this region in neural tissues. These results are preliminary, and a cautious interpretation is warranted as the probability is influenced by the modest power of both the GWAS[57].

insert Figure 3 here

Discussion

This is the first GWAS to investigate the genetic architecture of self-reported empathy. We identified three significant genetic correlations with the EQ and psychiatric

bioRxiv preprint doi: https://doi.org/10.1101/050682; this version posted November 14, 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.

conditions and psychological traits (anorexia nervosa, schizophrenia, and extraversion), providing insights into the shared genetic architecture. Although we did not identify any significant SNPs after correcting for multiple testing at $P < 2.5x10^{-8}$, we identified eleven SNPs of suggestive significance ($P < 1x10^{-6}$). Males and females perform differently on the tests, but there was limited evidence of a sex-specific genetic architecture.

We identified a significant negative genetic correlation between the EQ and autism. Several studies have identified lower self-reported empathy in individuals with autism, and our results mirror these studies[1,68]. This is likely to be driven by difficulties in understanding the mental states of others rather than responding to them. We also identified significant genetic correlations for the EQ with schizophrenia and anorexia nervosa. The empirical literature in general report deficits in cognitive empathy[4,70], but preserved or stronger affective empathy[6,70] and emotional contagion/personal distress[6] in individuals with schizophrenia compared to controls. Studies with anorexia nervosa, on the other hand, have yielded mixed results. Some studies suggest preserved empathy[71], some identify reduced cognitive empathy[72–74], and others identify greater emotional contagion/personal distress[75] in individuals with anorexia nervosa compared to controls. These studies are typically conducted in small samples, which may explain the different results in these heterogeneous conditions. Our results suggest that genetic variants associated with self-reported empathy slightly increase the risk for schizophrenia and anorexia nervosa; the latter remained significant even after using the females-only EQ dataset. A previous study[34] identified a significant genetic correlation between cognitive empathy (measured using the Eyes Test[76]) and anorexia nervosa, underscoring the importance of empathy as a genetic risk factor in anorexia nervosa. However, both cognitive empathy and anorexia nervosa are positively correlated with educational attainment[34,77], and it is possible that the correlation between cognitive empathy and anorexia nervosa may be mediated by educational attainment.

Here, self-reported empathy is not genetically correlated with educational attainment, which allows us to conclude that empathy indeed contributes to genetic risk for anorexia nervosa. Further, while cognitive empathy was not correlated with schizophrenia, self-reported empathy was positively and significantly correlated with schizophrenia, suggesting distinct roles for the two phenotypes in pathology. Schizophrenia and anorexia share significant positive genetic correlation ($r_g = 0.23 \pm 0.06$)[77], and it is possible that the pleiotropy between these two conditions may, in part, be mediated by genetic variants that contribute to empathy. This needs to be tested. The modest power of the remaining psychiatric GWAS

bioRxiv preprint doi: https://doi.org/10.1101/050682; this version posted November 14, 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.

precludes the identification of significant genetic correlations with self-reported empathy. Together with the GWAS on cognitive empathy[34], this study provides evidence for the distinct roles of different social processes in various psychiatric conditions.

Investigating genetic correlation results with psychological traits and measures of cognition further helped elucidate the genetic architecture of self-reported empathy. The EQ was significantly correlated with extraversion and nominally correlated with subjective wellbeing and conscientiousness. Both extraversion and conscientiousness correlate with empathy[65] which, in turn, contributes to subjective wellbeing[31]. Of the five personality factors, extraversion, conscientiousness and agreeableness have modest correlations with self-reported empathy[65]. We did not test for genetic correlation with agreeableness due to the low heritability of the trait. The direction of our genetic correlation results mirror observed phenotypic correlations and provide additional evidence for the positive role of self-reported empathy in subjective wellbeing.

This is also the first study to provide estimates of additive heritability explained by all the SNPs tested for self-reported empathy, and approximately 11% of variance was explained by SNPs. One study, investigating the heritability of the reduced EQ (18 items) in 250 twin pairs, identified a heritability of 0.32[12]. The literature on the heritability of empathy and prosociality is inconsistent, with heritability estimates ranging from 0.20[10] to 0.69[78], although a meta-analysis of different studies identified a heritability estimate of 0.35 (95% CI – 0.21 – 0.41)[79]. Our analysis therefore suggests that a third of the heritability can be attributed to common genetic variants. Like IQ[80], the heritability of empathy and prosociality behaviour changes with age[10]. We did not investigate the effect of age on heritability in our study.

We did not find any significant differences in heritability between males and females. Further, the male-female genetic correlations for both the traits were high. Despite the high genetic correlation, sex-specific correlations with anorexia nervosa were significant only for the females-only dataset. This suggests that the sex-specific genetic component of empathy can contribute differentially to psychiatric conditions. Several other factors may explain the observed phenotypic sex difference. For example, genetic variants for empathy may be enriched in sex-specific gene expression pathways. We conducted preliminary analysis by investigating if there is an enrichment in sex-differentially expressed genes in cortical tissue samples, but did not find significant enrichment. However, sex-specific gene expression is a dynamic process with both spatial and developmental differences[49,64]. Investigating across

bioRxiv preprint doi: https://doi.org/10.1101/050682; this version posted November 14, 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.

different tissues and developmental time points in well powered gene expression datasets will help better understand sex differences in empathy.

There are a few limitations that need to be taken into consideration in interpreting our results. The EQ is a self-report measure, and while it has excellent psychometric properties and construct validity, and it is unclear how much of the intrinsic biological variation in this trait is captured by it. Further, while this is the largest GWAS to date of self-reported empathy, it still is only modestly powered, reflected in our inability to identify genome-wide significant loci. This modest statistical power influences subsequent analysis, and we highlight several nominally significant results for further investigation in larger datasets.

In conclusion, the current study provides the first narrow sense heritability for empathy. While there is a highly significant difference on the EQ between males and females, heritability is similar, with a high genetic correlation between the sexes. We also identified significant genetic correlations between empathy and some psychiatric conditions and psychological traits, including autism. This global view of the genomic architecture of empathy will allow us to better understand psychiatric conditions, and improve our knowledge of the biological bases of neurodiversity in humans.

Acknowledgements

We thank Richard Bethlehem, Florina Uzefovsky, and Paula Smith 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 Templeton World Charity Foundation, 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. 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, Bethann S. Hromatka, Karen E. Huber,

bioRxiv preprint doi: https://doi.org/10.1101/050682; this version posted November 14, 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.

Aaron Kleinman, Nadia K. Litterman, 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. This work was supported by the National Human Genome Research Institute of the National Institutes of Health (grant number R44HG006981). 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.

References

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

2 Decety J, Bartal IB-A, Uzefovsky F, Knafo-Noam A. Empathy as a driver of prosocial behaviour: highly conserved neurobehavioural mechanisms across species. *Philos Trans R Soc London B Biol Sci* 2015; **371**: 20150077.

3 Derntl B, Seidel E-M, Schneider F, Habel U. How specific are emotional deficits? A comparison of empathic abilities in schizophrenia, bipolar and depressed patients. *Schizophr Res* 2012; **142**: 58–64.

4 Bora E, Gökçen S, Veznedaroglu B. Empathic abilities in people with schizophrenia.

bioRxiv preprint doi: https://doi.org/10.1101/050682; this version posted November 14, 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.

Psychiatry Res 2008; **160**: 23–9.

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

6 Lehmann A, Bahçesular K, Brockmann E-M, Biederbick S-E, Dziobek I, Gallinat J *et al.* Subjective experience of emotions and emotional empathy in paranoid schizophrenia. *Psychiatry Res* 2014; **220**: 825–33.

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

8 Thoma P, Schmidt T, Juckel G, Norra C, Suchan B. Nice or effective? Social problem solving strategies in patients with major depressive disorder. *Psychiatry Res* 2015; **228**: 835–42.

9 Baron-Cohen S. Autism: The Empathizing-Systemizing (E-S) Theory. *Ann N Y Acad Sci* 2009; **1156**: 68–80.

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

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

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

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

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

bioRxiv preprint doi: https://doi.org/10.1101/050682; this version posted November 14, 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.

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

16 Decety J. The neurodevelopment of empathy in humans. *Dev Neurosci* 2010; **32**: 257–67.

17 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.

18 Chapman E, Baron-Cohen S, Auyeung B, Knickmeyer R, Taylor K, Hackett G. Fetal testosterone and empathy: evidence from the Empathy Quotient (EQ) and the 'Reading the Mind in the Eyes' test. *Soc Neurosci* 2006; **1**: 135–48.

19 Siegal M, Varley R. Neural systems involved in 'theory of mind'. *Nat Rev Neurosci* 2002; **3**: 463–71.

20 Morelli SA, Rameson LT, Lieberman MD. The neural components of empathy: Predicting daily prosocial behavior. *Soc Cogn Affect Neurosci* 2014; **9**: 39–47.

21 Holgado Tello FP, Delgado Egido B, Carrasco Ortiz MA, Del Barrio Gandara M V. Interpersonal Reactivity Index: analysis of invariance and gender differences in spanish youths. *Child Psychiatry Hum Dev* 2013; **44**: 320–333.

22 Mestre MV, Samper P, Frías MD, Tur AM. Are women more empathetic than men? A longitudinal study in adolescence. *Span J Psychol* 2009; **12**: 76–83.

23 Connellan J, Baron-Cohen S, Wheelwright SJ, Batki A, Ahluwalia J. Sex differences in human neonatal social perception. *Infant Behav Dev* 2000; **23**: 113–118.

24 Christov-Moore L, Simpson EA, Coudé G, Grigaityte K, Iacoboni M, Ferrari PF. Empathy: gender effects in brain and behavior. *Neurosci Biobehav Rev* 2014; **46 Pt 4**: 604–27.

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

bioRxiv preprint doi: https://doi.org/10.1101/050682; this version posted November 14, 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.

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

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

28 Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K *et al.* Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am J Psychiatry* 2010; **167**: 748–51.

29 Lawrence EJ, Shaw P, Baker D, Baron-Cohen S, David AS. Measuring empathy: reliability and validity of the Empathy Quotient. *Psychol Med* 2004; **34**: 911–919.

30 Melchers MC, Montag C, Markett S, Reuter M. Assessment of empathy via self-report and behavioural paradigms: data on convergent and discriminant validity. *Cogn Neuropsychiatry* 2015; **20**: 157–71.

31 Bos EH, Snippe E, de Jonge P, Jeronimus BF, Duguid J, Bohlmeijer E. Preserving subjective wellbeing in the face of psychopathology: buffering effects of personal strengths and resources. *PLoS One* 2016; **11**: e0150867.

32 Spreng RN, McKinnon MC, Mar RA, Levine B. The Toronto Empathy Questionnaire: scale development and initial validation of a factor-analytic solution to multiple empathy measures. *J Pers Assess* 2009; **91**: 62–71.

33 Allison C, Baron-Cohen S, Wheelwright SJ, Stone MH, Muncer SJ. Psychometric analysis of the Empathy Quotient (EQ). *Pers Individ Dif* 2011; **51**: 829–835.

34 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 brain anatomy. *bioRxiv* 2016.

35 Baron-Cohen S, Bowen DC, Holt RJ, Allison C, Auyeung B, Lombardo M V. *et al.* The 'Reading the Mind in the Eyes' test: complete absence of typical sex difference in ~400 men and women with autism. *PLoS One* 2015; **10**: e0136521.

36 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*

bioRxiv preprint doi: https://doi.org/10.1101/050682; this version posted November 14, 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.

2011; **6**: e23473.

37 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.

38 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.

39 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.

40 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.

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

42 Hibar DP, Stein JL, Renteria ME, Arias-Vasquez A, Desrivières S, Jahanshad N *et al.* Common genetic variants influence human subcortical brain structures. *Nature* 2015; **520**: 224–9.

43 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.

44 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.

45 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.

46 Ge T, Chen C-Y, Neale BM, Sabuncu MR, Smoller JW, Tirrell LS. Phenome-wide heritability analysis of the UK Biobank. *PLOS Genet* 2017; **13**: e1006711.

bioRxiv preprint doi: https://doi.org/10.1101/050682; this version posted November 14, 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.

47 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.

48 Lek M, Karczewski KJ, Minikel E V., Samocha KE, Banks E, Fennell T *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016; **536**: 285–291.

49 Chen C-Y, Lopes-Ramos CM, Kuijjer ML, Paulson JN, Sonawane AR, Fagny M *et al.* Sexual dimorphism in gene expression and regulatory networks across human tissues. *bioRxiv* 2016.

50 Zheng J, Erzurumluoglu AM, Elsworth BL, Kemp JP, Howe L, Haycock PC *et al.* 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* 2016; **33**: 272–279.

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

52 Duncan LE, Yilmaz Z, Gaspar H, Walters RK, Goldstein JI, Anttila V *et al.* Significant locus and metabolic genetic correlations revealed in genome-wide association study of anorexia nervosa. *Am J Psychiatry* 2017; : appi.ajp.2017.1.

53 Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E *et al.* Discovery of the first genome-wide significant risk loci for ADHD. *bioRxiv* 2017.

54 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.

55 van den Berg SM, de Moor MHM, Verweij KJH, Krueger RF, Luciano M, Arias-Vasquez A *et al.* Meta-analysis of genome-wide association studies for extraversion: findings from the genetics of personality consortium. *Behav Genet* 2016; **46**: 170–82.

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

57 Pickrell JK, Berisa T, Liu JZ, Ségurel L, Tung JY, Hinds DA. Detection and

bioRxiv preprint doi: https://doi.org/10.1101/050682; this version posted November 14, 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.

interpretation of shared genetic influences on 42 human traits. *Nat Genet* 2016; **48**: 709–717.

58 Singh T, Walters JTR, Johnstone M, Curtis D, Suvisaari J, Torniainen M *et al.* Rare schizophrenia risk variants are enriched in genes shared with neurodevelopmental disorders. *bioRxiv* 2016.

59 Samocha KE, Robinson EB, Sanders SJ, Stevens C, Sabo A, McGrath LM *et al.* A framework for the interpretation of de novo mutation in human disease. *Nat Genet* 2014; **46**: 944–50.

60 Kosmicki JA, Samocha KE, Howrigan DP, Sanders SJ, Slowikowski K, Lek M *et al.* Refining the role of de novo protein-truncating variants in neurodevelopmental disorders by using population reference samples. *Nat Genet* 2017; **49**: 504–510.

61 McRae JF, Clayton S, Fitzgerald TW, Kaplanis J, Prigmore E, Rajan D *et al.* Prevalence and architecture of de novo mutations in developmental disorders. *Nature* 2017; **542**: 433–438.

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

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

64 Shi L, Zhang Z, Su B, Thompson PM, Thiel G. Sex biased gene expression profiling of human brains at major developmental stages. *Sci Rep* 2016; **6**: 21181.

65 Melchers MC, Li M, Haas BW, Reuter M, Bischoff L, Montag C. Similar personality patterns are associated with empathy in four different countries. *Front Psychol* 2016; **7**: 290.

66 Song YS, Lee H-J, Prosselkov P, Itohara S, Kim E. Trans-induced cis interaction in the tripartite NGL-1, netrin-G1 and LAR adhesion complex promotes development of excitatory synapses. *J Cell Sci* 2013; **126**: 4926–4938.

67 Seiradake E, Coles CH, Perestenko P V, Harlos K, McIlhinney RAJ, Aricescu AR *et al.* Structural basis for cell surface patterning through NetrinG-NGL interactions. *EMBO J* 2011; **30**: 4479–4488.

bioRxiv preprint doi: https://doi.org/10.1101/050682; this version posted November 14, 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.

68 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.

69 Peyrot WJ, Boomsma DI, Penninx BWJH, Wray NR. Disease and polygenic architecture: Avoid trio design and appropriately account for unscreened control subjects for common disease. *Am J Hum Genet* 2016; **98**: 382–391.

70 Horan WP, Reise SP, Kern RS, Lee J, Penn DL, Green MF. Structure and correlates of self-reported empathy in schizophrenia. *J Psychiatr Res* 2015; **66–67**: 60–6.

71 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.

72 Morris R, Bramham J, Smith E, Tchanturia K. Empathy and social functioning in anorexia nervosa before and after recovery. *Cogn Neuropsychiatry* 2014; **19**: 47–57.

73 Harrison A, Sullivan S, Tchanturia K, Treasure J. Emotion recognition and regulation in anorexia nervosa. *Clin Psychol Psychother*; **16**: 348–56.

74 Brewer R, Cook R, Cardi V, Treasure J, Bird G. Emotion recognition deficits in eating disorders are explained by co-occurring alexithymia. *R Soc open Sci* 2015; **2**: 140382.

75 Beadle JN, Paradiso S, Salerno A, McCormick LM. Alexithymia, emotional empathy, and self-regulation in anorexia nervosa. *Ann Clin Psychiatry* 2013; **25**: 107–20.

76 Baron-Cohen S, Wheelwright SJ, Hill J, Raste Y, Plumb I. The 'Reading the Mind in the Eyes' Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *J Child Psychol Psychiatry* 2001; **42**: 241–51.

77 Duncan LE, Yilmaz Z, Walters RK, Goldstein J, Anttila V, Bulik-Sullivan BK *et al.* Genome-wide association study reveals first locus for anorexia nervosa and metabolic correlations. *bioRxiv* 2016.

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

79 Knafo-Noam A, Uzefovsky F. Variation in empathy: The interplay of genetic and

bioRxiv preprint doi: https://doi.org/10.1101/050682; this version posted November 14, 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.

environmental factors. In: Legerstee M, Haley DW, Bornstein MH (eds). *The infant mind: Origins of the social Brain*. The Guilford Press: New York, 2013, pp 97–121.

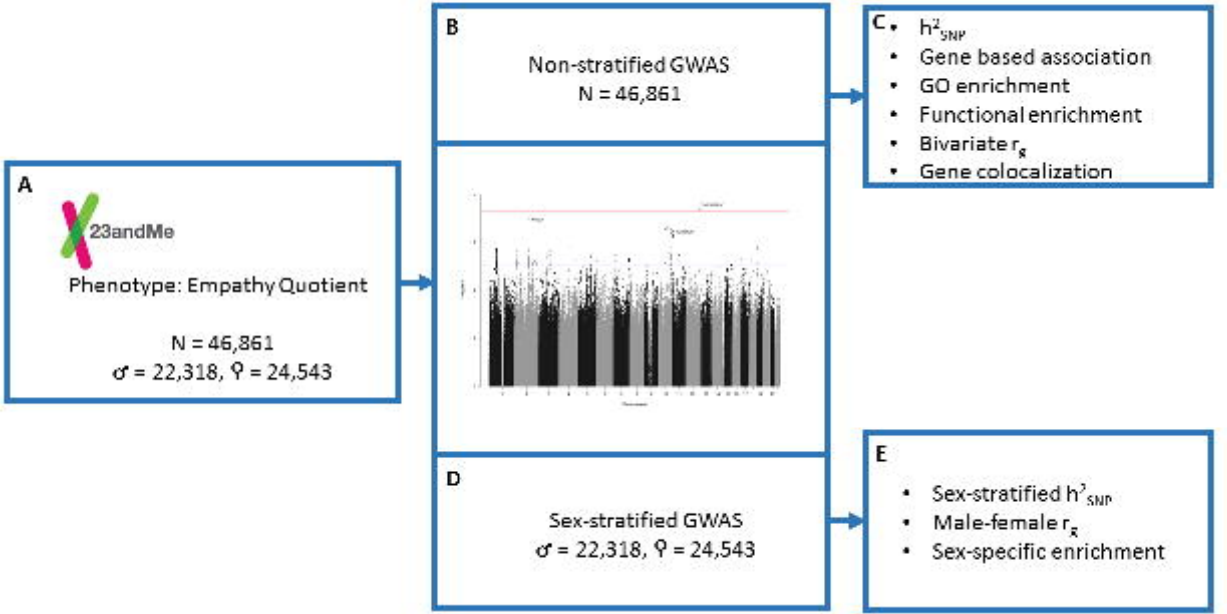
80 Bouchard TJ. The Wilson Effect: the increase in heritability of IQ with age. *Twin Res Hum Genet* 2013; **16**: 923–30.

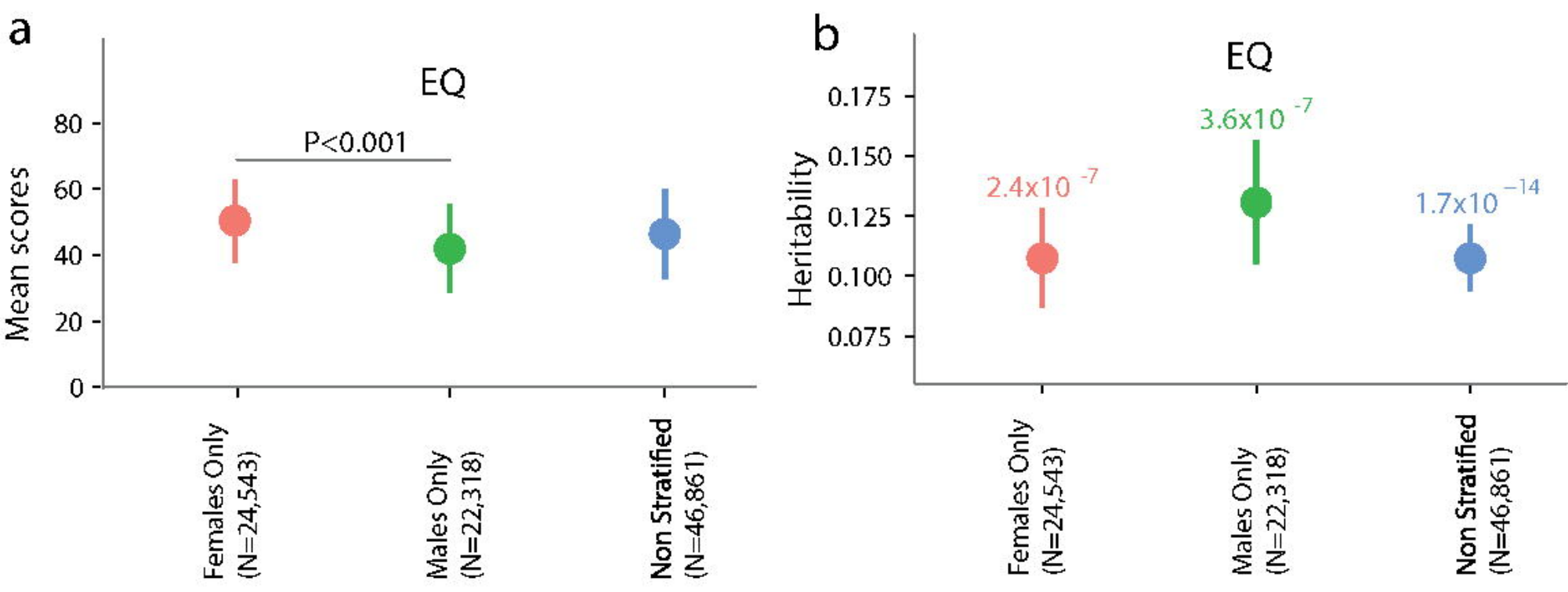
Figure Legends

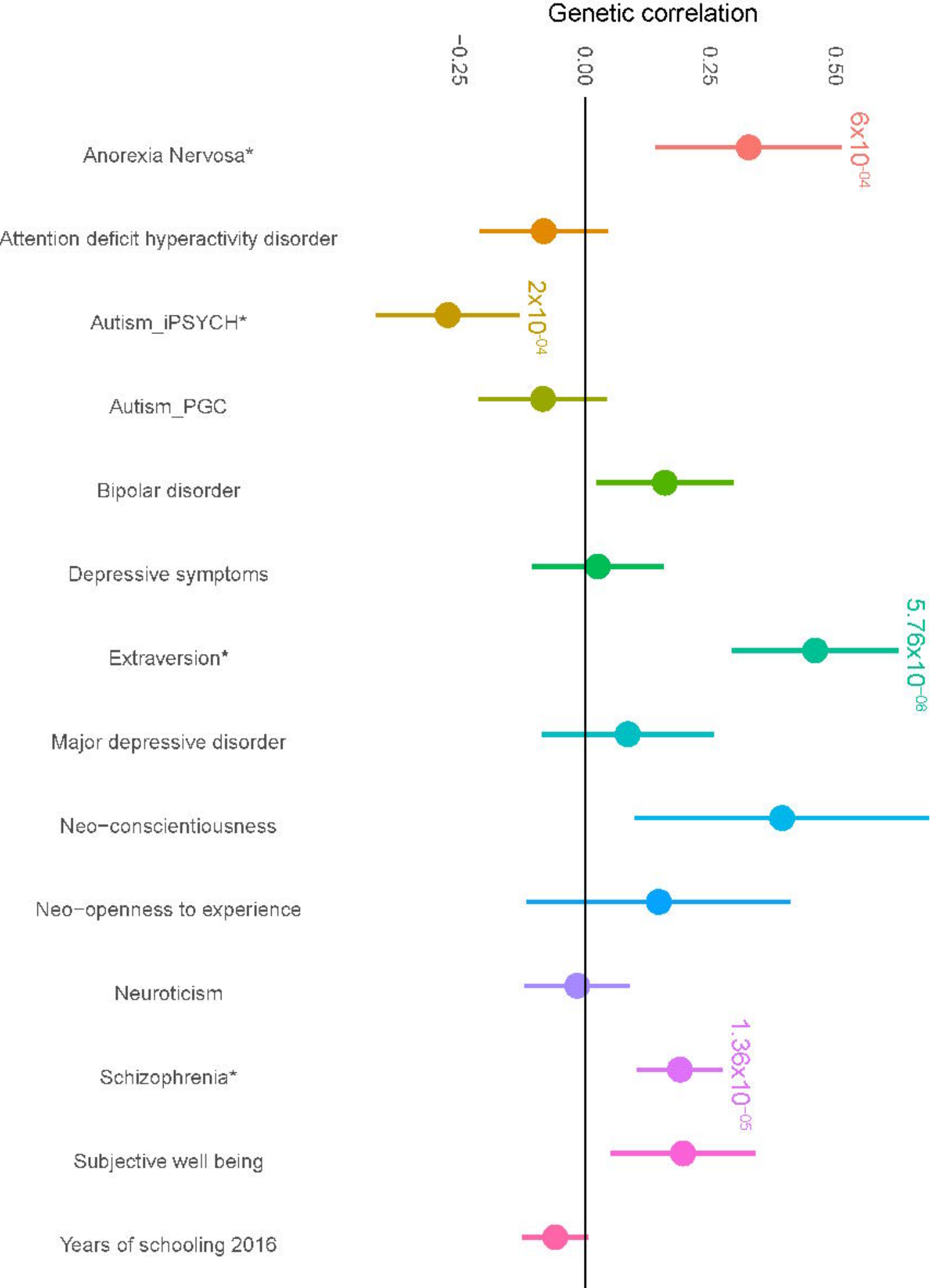
Figure 1: Schematic diagram of the study protocol. *Phenotyping was conducted in research participants from 23andMe., Inc using the 60 question Empathy Quotient (Panel A). Three GWAS analyses were conducted: non-stratified GWAS (Panel B) and sex-stratified GWAS (Panel D) in unrelated individuals of primarily European ancestry. Summary non-stratified GWAS data was used to conduct SNP heritability, gene, pathway, and functional enrichment, genetic correlations with psychiatric conditions, psychological traits and education, and Bayesian gene colocalization (Panel C). Summary sex-stratified GWAS data was used to conduct sex-specific SNP heritability, genetic correlation between the male and female datasets, and enrichment in sex-differentially expressed genes (Panel E).*

Figure 2: Mean scores and heritability estimates for the EQ. *Mean scores and standard deviations for scores on the EQ (a). The effect size of the difference between males and females for the EQ scores was Cohen's d= 0.65. The bar on top and the number represents the statistical significance of the male-female difference in mean scores for the EQ. Mean estimates and standard errors for heritability for scores on the EQ (b) for the females-only GWAS, males-only GWAS and the non-stratified GWAS. Numbers on top of the graphs represent P-values for each heritability estimate. Note, the y-axis does not start at 0.*

Figure 3: Genetic correlations between the EQ and other conditions. *Mean and 95% confidence intervals shown for genetic correlations between empathy and other conditions. P-values are provided significant correlations after Bonferroni correction.*


A
23andMe
Phenotype: Empathy Quotient
N = 46,861
♂ = 22,318, ♀ = 24,543
B
Non-stratified GWAS
N = 46,861
C
• h^2_{SNP}
• Gene based association
• GO enrichment
• Functional enrichment
• Bivariate r_g
• Gene colocalization
D
Sex-stratified GWAS
♂ = 22,318, ♀ = 24,543
E
• Sex-stratified h^2_{SNP}
• Male-female r_g
• Sex-specific enrichment


a
EQ
P<0.001
Mean scores
0
20
40
60
80
Females Only (N=24,543)
Males Only (N=22,318)
Non Stratified (N=46,861)
b
EQ
Heritability
0.075
0.100
0.125
0.150
0.175
2.4x10 -7
3.6x10 -7
1.7x10 −14
Females Only (N=24,543)
Males Only (N=22,318)
Non Stratified (N=46,861)


Genetic correlation
−0.25
0.00
0.25
0.50
Anorexia Nervosa*
6x10-04
Attention deficit hyperactivity disorder
Autism_iPSYCH*
2x10-04
Autism_PGC
Bipolar disorder
Depressive symptoms
Extraversion*
5.76x10-08
Major depressive disorder
Neo−conscientiousness
Neo−openness to experience
Neuroticism
Schizophrenia*
1.36x10-05
Subjective well being
Years of schooling 2016