Genome-wide association study of social relationship satisfaction:

significant loci and correlations with psychiatric conditions

Varun Warrier[1,#], Thomas Bourgeron[2,3,4,*], Simon Baron-Cohen[1,5,*,#]

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. CLASS Clinic, Cambridgeshire and Peterborough NHS Foundation Trust (CPFT), Cambridgeshire, United Kingdom

* Co-senior authors

#Correspondence to : Varun Warrier – vw260@medschl.cam.ac.uk, Simon Baron-Cohen

– sb205@cam.ac.uk

Abstract

Dissatisfaction in social relationships is reported widely across many psychiatric conditions. We investigated the genetic architecture of family relationship satisfaction and friendship satisfaction in the UK Biobank. We leveraged the high genetic correlation between the two phenotypes (r_g = 0.87±0.03; $P < 2.2x10^{-16}$) to conduct multi-trait analysis of Genome Wide Association Study (GWAS) ($N_{effective}$ family = 164,112; $N_{effective}$ friendship = 158,116). We identified two genome-wide significant associations for both the phenotypes: rs1483617 on chromosome 3 and rs2189373 on chromosome 6, a region previously implicated in schizophrenia. eQTL and chromosome conformation capture in neural tissues prioritizes several genes including *NLGN1*. Gene-based association studies identified several significant genes, with highest expression in brain tissues. Genetic correlation analysis identified significant negative correlations for multiple psychiatric conditions including highly significant negative correlation with cross-psychiatric disorder GWAS, underscoring the central role of social relationship dissatisfaction in psychiatric diagnosis. The two phenotypes were enriched for genes that are loss of function intolerant. Both phenotypes had modest, significant additive SNP heritability of approximately 6%. Our results underscore the central role of social relationship satisfaction in mental health and identify genes and tissues associated with it.

Introduction

Social relationship satisfaction is one contributor to subjective wellbeing, alongside other domains such as occupational, financial, and health satisfaction[6]. Difficulties in forming and maintaining social relationships are reported widely in psychiatric conditions[1-4]. Positive social relationship satisfaction can both reduce the risk and ameliorate the severity of a

psychiatric condition. In the DSM-5[5], difficulties in social functioning is one of the criteria for diagnosing conditions such as autism, anorexia nervosa, schizophrenia, and bipolar disorder. However, little is known about the genetic architecture of social relationship satisfaction, and if social relationship dissatisfaction genetically contributes to risk for psychiatric conditions. To our knowledge, there is no twin study that has focussed on twin heritability of social relationship satisfaction. A few studies have investigated the heritability of subjective wellbeing and have identified heritability estimates between 38 – 50%[7,8].

Here we report on two genome-wide association studies (GWAS) of social relationship satisfaction in the UK Biobank. We focus on friendship satisfaction and family relationship satisfaction, measured using two questionnaires. We leverage the high genetic correlation between the two phenotypes to: 1. Identify the genetic architecture of friendship and family relationship satisfaction; 2. Identify the genetic correlation between these two phenotypes and specific psychiatric conditions and psychological phenotypes; 3. Prioritize associated genes, and enriched gene sets and tissues for the two phenotypes; 4. Quantify cross-phenotype polygenic predictive power of the two phenotypes.

We identified two genetic loci associated with the phenotypes. We further identified significant genetic correlations with several psychiatric conditions, and demonstrated an enrichment for brain tissues, and conclude by prioritizing genes and gene sets for further investigation.

Methods

Phenotypes and participants

3

Participants were individuals from the UK Biobank. Participants were asked "In general, how satisfied are you with your family relationships?" and "In general, how satisfied are you with your friendships?" Participants could choose one of eight options: "Extremely happy", "Very happy", "Moderately happy", "Moderately unhappy", "Very unhappy", "Extremely unhappy", "Do not know", and "Prefer not to answer". We excluded individuals who responded "Do not know", or "Prefer not to answer" The phenotypes were coded in three waves. We combined responses from three waves, and removed participants who had responded in more than one wave (N = 178,675 for family and N = 178,721 for friendship). Responses were recoded from 1 to 6 with 1 being "Extremely unhappy", and 6 being "Extremely happy". We removed participants who were not genotyped, who were not of "British ancestry", who were outliers for heterozygosity, and whose reported sex did not match their genetic sex. This left us with N = 139,603 for family, and N = 139,826 for friendship. Finally, we removed related individuals from this list resulting in a total of N = 134,681 (family) and N = 134,941 (friendship) unrelated individuals for further analyses. A total of 131,790 individuals were included in both the analyses. At the time of recruitment, participants were between 40 – 69 years of age. All participants provided informed consent to participate in the UK Biobank. In addition, we obtained ethical approval from the University of Cambridge Human Biology Research Ethics Committee (HBREC) to use de-identified data from the UK Biobank and ALSPAC for this study. ALSPAC is Avon Longitudinal Study of Parents and Children[27] which was used in a polygenic risk score analysis (see below).

Genetic analyses

Details of the UK Biobank genotyping, imputation, and quality control procedures is available elsewhere[9]. Genetic association was conducted using Plink 2.0[10] (https://www.cog-

genomics.org/plink/2.0/). We included sex, year of birth, genotyping batch and the first 40 genetic principal components as covariates. We excluded SNPs that failed Hardy-Weinberg Equilibrium (P < 1x10^{-6}), had a minor allele frequency < 0.01, and had a per-SNP genotyping rate < 95%, and imputation INFO < 0.1. We further excluded SNPs not in the Haplotype Reference Consortium (http://www.haplotype-reference-consortium.org/)[11]. We excluded individuals who were ancestry outliers, and had per-individual genotyping rate < 90%.

To increase the effective sample size, we leveraged the high genetic correlation between the two phenotypes to conduct MTAG[12] (or multi-trait analysis of GWAS) (https://github.com/omeed-maghzian/mtag/). MTAG is an extension of the standard inverse variance weighted meta-analysis that considers the genetic correlation between phenotypes to estimate phenotype-specific effects. Our phenotypes are excellent for MTAG given both the high genetic correlation between the two phenotypes and the similar mean chi-square. We do not expect the estimates to be biased.

Clumping of the independent SNPs was conducted using Plink[10], using an r^2 of 0.6. *Winner's curse correction* was conducted using FIQT, which is an FDR based inverse quantile transformation[13].

Genetic correlations were conducted using LD score regression (LDSR)[14,15] (https://github.com/bulik/ldsc/wiki). We used north-west European population LD scores, did not constrain the intercept and used the pre-MTAG summary statistics to obtain unbiased genetic correlations. We ran correlations with psychiatric and psychological phenotypes **(Supplementary Table 1)**. Summary scores were obtained from the PGC (http://www.med.unc.edu/pgc/results-and-downloads), the SSGAC

(https://www.thessgac.org/), and the CTG (https://ctg.cncr.nl/software/summary_statistics). In addition, summary statistics for systemizing, self-reported empathy[16], and cognitive empathy[17] were obtained from 23andMe, Inc. Summary statistics for the iPsych autism replication dataset were obtained from the iPsych autism team. Given the correlation between the phenotypes, we identified significant genetic correlations using a Benjamini-Hochberg FDR < 0.05, taking into account all the tests conducted for both the phenotypes combined.

Heritability analyses were conducted using both LDSR (all participants)[14] and GCTA-GREML[18] (for computational efficiency, this was conducted on one-fifth of the total participants, N = 26,000). To keep the methods as identical as possible, GCTA heritability was calculated after including sex, age, batch and the first 40 genetic principal components as covariates. Bivariate genetic correlations between the two phenotypes were conducted using both LDSR and GCTA.

Functional annotation was performed on the MTAG-GWAS datasets. We used FUMA[19] (http://fuma.ctglab.nl/) to identify eQTLs and chromatin interactions for the genome-wide significant results. eQTLS were identified using data from BRAINEAC and GTEx[20] brain tissues. Chromatin interactions were identified using Hi-C data from the dorsolateral prefrontal cortex, hippocampus, and neural progenitor cells. Significant mapping was identified at a Benjamini-Hochberg FDR < 0.05. Gene-based association analyses was conducted using MAGMA[21] within FUMA, and significant genes were identified using a Bonferroni-corrected threshold < 0.05 for each phenotype, as provided within FUMA.

We conducted partitioned heritability[22] analyses using the MTAG-GWAS dataset using the baseline categories and additional categories: genes that are loss-of-function

intolerant[23], and genes with brain specific annotation. Significant enrichments were identified using a Benjamini-Hochberg FDR < 0.05 applied to both the phenotypes combined.

To identify phenotype relevant tissues based on gene expression, we conducted partitioned heritability analyses applied to tissue-specific genes[24]. We focussed on GTEx consortium based gene expression from 13 brain regions (Cortex, Anterior cingulate cortex, frontal cortex, cerebellum, cerebellar hemisphere, putamen, nucleus accumbens, caudate, substantia nigra, hippocampus, spinal cord, amygdala, and hypothalamus). Results were significant at a Benjamini-Hochberg FDR < 0.05. In addition, gene expression analyses for general tissue types and specific tissue types were also conducted using FUMA. These two methods are not identical. In partitioned heritability, the top 10% of gene with tissue specific expression (focal brain region vs non-focal brain region) were included with a 100kb window on either side of the transcribed region. In FUMA, all genes are tested using a linear regression framework. Here, the expression of genes in a particular tissue is regressed against the gene Z-values with the average expression of the gene in all tissues included alongside usual covariates (gene length, SNP density etc). Thus, while FUMA tests all the genes, partitioned heritability tests only the top decile.

Cell specific enrichment[25] was conducted for three cell types (Neurons, astrocytes and oligodendrocytes) using both LDSR partitioned heritability and MAGMA gene set enrichment. Results were FDR corrected (P < 0.05). For the gene-set enrichment with MAGMA, we used the top 500 cell-type specific genes for the three cell types. For oligodendrocytes, we used clustered both newly formed and myelinating oligodendrocytes into one group. The cell-specific genes can be downloaded here: https://web.stanford.edu/group/barres_lab/brain_rnaseq.html .

7

Polygenic regression analyses using the MTAG-GWAS results were conducted using PRSice[26] to generate polygenic scores. Regression analysis was conducted in R version 3.3.3. Polygenic scoring was conducted in approximately 5,600 phenotyped and unrelated individuals from the Avon Longitudinal Study of Parents and Children[27]. We considered total scores from three parent-reported child-based questionnaires, collected between the ages of 7 and 10: the Children's Communication Checklist (CCC)[28], the Social and Communication Disorder Checklist[29], and the Strengths and Difficulties Questionnaire (SDQ)[30]. These three questionnaires are widely used to assess difficulties in child development. In addition, we also tested five subdomains of the SDQ (prosocial behaviour, hyperactivity, emotional difficulties, peer difficulties and conduct problems). Prosocial behaviour and CCC were reverse coded to indicate difficulties in prosocial behaviour and communication. Polygenic score regression was conducted using a negative binomial model (MASS package in R: https://cran.r-project.org/web/packages/MASS/MASS.pdf) as this provided the best fit. Sex and the first two genetic principal components were included as covariates. Incremental variance explained due to polygenic scores was calculated using Nagelkerke's pseudo-R^2. Further details of the ALSPAC cohort are provided in the **Supplementary Note**.

Results

Phenotypic distributions

Both phenotypes had a unimodal, near-normal distribution (**Supplementary Figure 1**). The Spearman's rank correlation between friendship and family relationship satisfaction was only 0.52 (P < 2.2x10^{-16}) (**Supplementary Figure 2**), suggesting that they are not identical phenotypes. This allowed us to investigate the genetic architecture of the two

8

phenotypes separately. Both sex and age were significantly associated with the two phenotypes, though the effect was small (**Supplementary Figures 3 and 4**).

Genetic correlation

Despite the modest phenotypic correlation, genetic correlation between family relationship satisfaction and friendship satisfaction was high ($r_g = 0.87\pm0.03$; $P < 2.2\mathrm{x}10^{-16}$), suggesting a similar genetic architecture between the two phenotypes. We observed a similarly high bivariate genetic correlation using GCTA REML ($r_g = 1 \pm 0.11$; $P < 2.2\mathrm{x}10^{-16}$) in our subsample of participants. This similarity was reflected in the largely similar genetic correlation between the two phenotypes and other psychiatric and psychological phenotypes. After FDR correction, we identified significant negative genetic correlations between family relationship satisfaction and six of the seven psychiatric conditions tested (anxiety, autism, anorexia nervosa, bipolar disorder, depression, and schizophrenia) and between friendship satisfaction and four of the seven psychiatric conditions (autism, bipolar disorder, depression, and schizophrenia). We replicated the negative genetic correlations with autism using a separate autism GWAS dataset (autism-iPsych). Underscoring the contribution to the two phenotypes to most psychiatric conditions, we also identified significant negative genetic correlation for both phenotypes and cross-disorder psychiatric GWAS. For cognitive and psychological phenotypes, we identified negative genetic correlations for both phenotypes and educational attainment, cognitive aptitude, depressive symptoms and neuroticism. In contrast, we identified significant positive genetic correlations between both phenotypes and conscientiousness, empathy, extraversion, and subjective wellbeing. Genetic correlations are provided in **Figure 1 and Supplementary Table 1**.

Insert Figure 1 here

SNP heritability

LDSR based SNP-heritability was 0.065 (0.005) for friendship satisfaction and 0.066 (0.005) for family relationship satisfaction. GCTA based heritability for a subset of the participants identified a significant heritability of 0.053 (0.014) for friendship satisfaction and 0.056 (0.014) for family relationship satisfaction (**Supplementary Figure 5 and Supplementary Table 2**).

Genetic association analyses

Genome-wide association analyses did not identify any significant results. Manhattan and QQ-plots are provided in **Supplementary Figure 6**. The λ_{GC} was 1.15 for both phenotypes. However, the LDSR intercept was 1.0018 (0.007) for family relationship satisfaction and 1.0002 (0.0085) for friendship satisfaction, suggesting negligible inflation due to population stratification. The high genetic correlation between the two phenotypes was supported by concordant effect direction for all the 13 independent loci with $P < 1x10^{-6}$ for the two phenotypes (**Supplementary Figure 7**). Six of these had a significant P-values in the other phenotype after Bonferroni correction (**Supplementary Figure 7 and Supplementary Table 3**).

Given the similar mean chi-square of the two phenotypes (1.13 for family relationship satisfaction and 1.14 for friendship satisfaction) and the high genetic correlation, we conducted a modified inverse variance weighted meta-analysis (MTAG). This increased the effective sample size to 164,112 for family relationship satisfaction and over 158,166 for friendship satisfaction. There was a high genetic correlation between the pre-MTAG and post-MTAG GWAS datasets (Family: 0.98±0.005, $P < 2.2x10^{-16}$; Friendship: $r_g =$

0.98 ± 0.004; P < 2.2×10^{-16}). We identified two significant loci associated with both the phenotypes **(Supplementary Table 4)**. On Chromosome 3, the lead SNP was rs1483617 for both phenotypes (P = 1.6×10^{-8} for family relationship satisfaction and 4.55×10^{-8} for friendship satisfaction). In addition, on chromosome 6, we identified another locus with rs2189373 as the lead SNP for both phenotypes (P = 1.06×10^{-8} for family and P = 4.21×10^{-8} for friendship). The top-SNPs explained 0.019% of the variance, which reduced to 0.0027% after correcting for winner's curse. Manhattan and QQ-plots for the GWAS are provided in **Figure 2.**

Insert Figure 2 here

Functional annotation

We investigated genes associated with the two loci using eQTL and chromosome interaction analysis in brain tissues **(Supplementary Tables 5 and 6)**. The analyses prioritized several genes including *NLGN1* which is involved in the postsynaptic complex in excitatory synapses (rs1483617) and interacts with *NRXN1*, a gene implicated in schizophrenia[31,32], autism[33], and intellectual disability[34]. A second gene identified, *KCNMB2,* modulates the calcium sensitivity in the BK channels. The second locus on chromosome 6 (rs2189373) is in the MHC complex. It is an intergenic SNP in *HCG17*, a gene which has been implicated in schizophrenia[35]. SNPs in high LD ($r^2 > 0.8$) with the lead SNP have been implicated in schizophrenia (rs2021722, P = 2.2×10^{-12})[36] (rs2523722 , P = 1.47×10^{-16})[37] and in the cross-disorder psychiatric GWAS (rs2517614, P = 8.9×10^{-7})[38]. When comparing the two previous studies and our current study, the effect allele which increased the risk for the psychiatric condition is in the same haplotype as the effect allele which decreased social relationship satisfaction. We note here that this region was not genotyped in the latest GWAS of schizophrenia[39], and no SNP in $r^2 > 0.8$ with rs2189373 was available in

this schizophrenia summary statistics. eQTL and CCC identified several genes in this locus. Notably, one of the genes identified (*ZNF184*) was also prioritized in a recent large-scale GWAS on neuroticism[40]. **Figure 3** provides the circos plots for the two loci. **Supplementary Figures 8 and 9** provide the local LD plots for the two phenotypes

Insert Figure 3 here

Using two different methods (FUMA and LDSR-partitioned heritability), we identified high enrichment in brain tissues including the pituitary, pointing to the significant role of neural tissues for both phenotypes **(Supplementary Figures 10 and 11, Supplementary Tables 7 - 9)**. Brain specific chromatin modifications showed a significant, 3-fold enrichment using LDSR, accounting for nearly a third of the heritability **(Supplementary Tables 7 and 8)**. In addition, both methods identified significant enrichment for genes with PLI > 0.9 (Family: Partitioned-h^2 P = 6.81×10^{-7}, MAGMA P = 3.06×10^{-4}; Friendship: Partitioned-h^2 P = 1.56×10^{-6}; MAGMA_P = 9.62×10^{-5}) **(Supplementary Tables 7 and 8)**. Baseline partitioned heritability also identified enrichment for conserved regions and H3K4me1 modifications for both phenotypes.

We investigated enrichment in specific brain regions using FUMA and partitioned heritability **(Supplementary Figures 10 and 11, Supplementary Table 9)**. FUMA identified a significant enrichment for the cerebellum. In contrast, partitioned heritability identified a significant enrichment for the cortex and the anterior cingulate cortex. This is likely to be due to methodological differences. Cell type specific analysis did not identify a significant enrichment for any specific brain types, using both partitioned heritability and MAGMA **(Supplementary Tables 10 and 11)**.

Gene-based analyses using MAGMA identified thirteen significant genes for family relationship satisfaction and 12 significant genes for friendship satisfaction (**Supplementary Tables 12 and 13)**. The top gene in both the phenotypes was *SHISA5*, a gene that encodes a transmembrane protein on the endoplasmic reticulum.

Polygenic score analyses

We identified a high correlation between polygenic scores for the two phenotypes in the ALSPAC cohort (r = 0.93; P < 2.2×10^{-16}), which reflects the high genetic correlation between the traits **(Supplementary Figure 12)**. Polygenic prediction analyses were conducted for three childhood questionnaires and 5 subscales (Methods). After FDR correction, polygenic scores did not significantly predict variance in any of the phenotypes **(Supplementary Table 14)**. The variance explained was small, with the highest variance explained being for childhood prosocial behaviour **(Supplementary Table 14 and Supplementary Figure 13 and 14)**.

Discussion

We present the results of a large-scale genome-wide association study of social relationship satisfaction in the UK Biobank measured using family relationship satisfaction and friendship satisfaction. Despite the modest phenotypic correlations, there was a significant and high genetic correlation between the two phenotypes, suggesting a similar genetic architecture between the two phenotypes.

13

We first investigated if the two phenotypes were genetically correlated with psychiatric conditions. As predicted, most if not all psychiatric conditions had a significant negative correlation for the two phenotypes. We replicated the correlation with autism using a second dataset. We observed significant negative genetic correlation between the two phenotypes and a large cross-condition psychiatric GWAS[38]. This underscores the importance of social relationship dissatisfaction in psychiatric conditions. The genetic correlations identified here are very similar to those identified between psychiatric conditions, psychological phenotypes and subjective well-being[41]. One notable exception is the negative genetic correlation between measures of cognition[42,43] and the two phenotypes. Whilst subjective wellbeing is positively genetically correlated with measures of cognition, we identify a small but statistically significant negative correlation between measures of correlation and the two phenotypes. A recent study highlighted that people with very high IQ scores tend to report lower satisfaction with life with more frequent socialization[44]. This also highlights the distinctions between subjective wellbeing and social relationship satisfaction. Subjective wellbeing encompasses several aspects of wellbeing including social relationships satisfaction.

We leveraged the high genetic correlation between the two phenotypes to identify significant genes using a method known as MTAG[12]. MTAG makes several assumptions, the key assumption being that the shared variance-covariance matrix is uniform across all SNPs. Despite this, we were comfortable applying MTAG to the two phenotypes due to both the high genetic correlation measured using two different methods, and the similar statistical power of the two GWAS, identified using the similar mean chi-square.

14

Our MTAG GWAS identified two significant loci. The first (rs1483617) is, as far as we are aware, a novel locus that has not been implicated previously in any GWAS of a neuropsychiatric condition. eQTL and chromosome interactions prioritize several genes that interact with this locus including *NLGN1*, which codes for the post-synaptic protein neuroligin1. NLGNs along with NRXNs and SHANKs are integral component of the synaptic complex, and mutations in these group of proteins have been implicated in several neuropsychiatric conditions including autism[45]. NLGN1 is located in the post-synaptic membrane of glutamatergic synapses, and can modulate the development of glutamatergic synapses in an activity dependent manner[45]. Another gene prioritized at this locus, *KCNMB2,* encodes a protein that is a component of the BK channels which are large conductance potassium channels that are sensitive to both intracellular voltage changes and changes in calcium ions.

The second locus (rs2189373) is in the MHC, a region of complex LD structure. This locus has been implicated previously in schizophrenia, and is nominally associated with the cross-condition psychiatric GWAS. Interestingly, this locus was not investigated in the recent large-scale GWAS of schizophrenia[39], as inferred from the summary statistics available. The effect direction of the SNPs when considering the haplotype structure of this region align with the genetic correlation between social relationship satisfaction and psychiatric risk. In other words, alleles that increase the risk for schizophrenia are in the same haplotype as alleles that decrease friendship satisfaction. The functional consequences of this locus must be formally tested.

Multiple different lines of evidence also point to the central role of brain tissues for the two phenotypes. First, using partitioned heritability, we identified a significant, threefold-

15

enrichment in brain specific annotations for both phenotypes, accounting for about 30% of the total heritability. This mirrored the high enrichment in brain specific expression identified by FUMA, including the pituitary. We also sought to identify where within the brain tissue GWAS signals for the two phenotypes are enriched in. Here, FUMA and partitioned heritability identified enrichment in different tissues. While FUMA prioritized the cerebellum, heritability prioritized the cortex. The discordance in results is due to the different definition of tissue specific expression used by the two methods. While partitioned heritability compares focal brain region expression against other brain regions, FUMA compares the focal brain region's expression against all tissue types. Hence, partitioned heritability is more likely to tag brain region specific expression patterns.

In addition to enrichment for brain tissues, we also identified an enrichment for genes that are intolerant to loss of function mutations[23]. A fifth of the heritability can be attributed to pLI > 0.9 genes. We utilized pLI quantified from the non-psychiatric population cohort for this analysis given the considerable genetic correlation between the two phenotypes and various psychiatric conditions. Loss of function mutations in these genes lead to severe biochemical consequences, and are implicated in several neuropsychiatric conditions. For example, *de novo* loss of function mutations in pLI intolerant genes confers significant risk for autism[46]. Our results suggest that pLI > 0.9 genes contribute to psychiatric risk through both common and rare genetic variation.

Whilst we were unable to investigate the polygenic prediction in other samples with same phenotypic measures, we investigated polygenic prediction across a range of different phenotypes in children. We chose three specific questionnaires that have been associated with various psychiatric conditions[47–49], and subdomains in the SDQ. After FDR correction, we

did not identify any significant polygenic prediction for the two phenotypes. The variance explained was small. This is due to three reasons – the modest genetic correlation between social relationship satisfaction and the childhood phenotypes, the modest statistical power of this study due to both low per-SNP variance explained and low additive heritability, and the relatively small sample size of the ALSPAC cohort. Encouragingly, the phenotype for which the variance explained was the highest is childhood prosocial behaviour.

In conclusion, we identify two significant loci association with social relationship satisfaction, one of which has been previously implicated in schizophrenia. Comprehensive functional annotation prioritizes specific genes for functional follow-up, and identifies tissue specific enrichment pattern. Genetic correlation analyses highlight the importance of social relationship dissatisfaction in psychiatric conditions, and enrichment in loss-of-function intolerance genes suggest a degree of overlap between rare and common variants in a set of genes for psychiatric conditions. Social relationship dissatisfaction is not just a consequence of having a psychiatric condition because both phenotypes have genetic correlates, meaning they are both valid phenotypes. Future research should test the function of the two significant SNPs (rs1483617 and rs2189373) we identified, and their role in each psychiatric condition.

References

1.   Cable, N., Bartley, M., Chandola, T. & Sacker, A. Friends are equally important to men and women, but family matters more for men's well-being. *J. Epidemiol. Community Health* **67,** 166–171 (2013).

2.   Teo, A. R., Choi, H., Valenstein, M., Haan, M. de & Dyck, R. van. Social

Relationships and Depression: Ten-Year Follow-Up from a Nationally Representative Study. *PLoS One* **8,** e62396 (2013).

3. Stevens, A. K., McNichol, J. & Magalhaes, L. Social relationships in schizophrenia: A review. *Personal. Ment. Health* **3,** 203–216 (2009).

4. Greenberg, S., Rosenblum, K. L., McInnis, M. G. & Muzik, M. The role of social relationships in bipolar disorder: A review. *Psychiatry Res.* **219,** 248–254 (2014).

5. American Psychiatric Association. *The Diagnostic and Statistical Manual (5th ed.).* (2013).

6. Diener, E., Suh, E. M., Lucas, R. E. & Smith, H. L. Subjective well-being: Three decades of progress. **125,** 276–302 (1999).

7. Stubbe, J. H., Posthuma, D., Boomsma, D. I. & De Geus, E. J. C. Heritability of life satisfaction in adults: a twin-family study. *Psychol. Med.* **35,** 1581 (2005).

8. Lykken, D. & Tellegen, A. Happiness Is a Stochastic Phenomenon. *Psychol. Sci.* **7,** 186–189 (1996).

9. Bycroft, C. *et al.* Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRxiv* (2017).

10. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81,** 559–75 (2007).

11. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48,** 1279–1283 (2016).

12. Turley, P. *et al.* MTAG: Multi-Trait Analysis of GWAS. *bioRxiv* (2017).

13. Bigdeli, T. B. *et al.* A simple yet accurate correction for winner's curse can predict signals discovered in much larger genome scans. *Bioinformatics* **32,** 2598–2603 (2016).

14. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from

polygenicity in genome-wide association studies. *Nat. Genet.* **47,** 291–295 (2015).

15. Bulik-Sullivan, B. *et al. An Atlas of Genetic Correlations across Human Diseases and Traits. bioRxiv* (Cold Spring Harbor Labs Journals, 2015). doi:10.1101/014498

16. Warrier, V. *et al. Genome-wide analyses of empathy and systemizing: heritability and correlates with sex, education, and psychiatric risk. bioRxiv* (Cold Spring Harbor Labs Journals, 2016). doi:10.1101/050682

17. Warrier, V. *et al.* Genome-wide meta-analysis of cognitive empathy: heritability, and correlates with sex, neuropsychiatric conditions and cognition. *Mol. Psychiatry* (2017). doi:10.1038/mp.2017.122

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

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

20. Ardlie, K. G. *et al.* Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science (80-.).* **348,** 648–60 (2015).

21. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11,** 1–19 (2015).

22. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47,** 1228–1235 (2015).

23. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536,** 285–291 (2016).

24. Finucane, H. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *bioRxiv* (2017).

25. Zhang, Y. *et al.* An RNA-Sequencing Transcriptome and Splicing Database of Glia, Neurons, and Vascular Cells of the Cerebral Cortex. *J. Neurosci.* **34,** (2014).

26. Euesden, J., Lewis, C. M. & O'Reilly, P. F. PRSice: Polygenic Risk Score software. *Bioinformatics* **31,** 1466–1468 (2015).

27. Boyd, A. *et al.* Cohort Profile: The 'Children of the 90s'—the index offspring of the Avon Longitudinal Study of Parents and Children. *Int. J. Epidemiol.* **42,** 111–127 (2013).

28. Bishop, D. V. Development of the Children's Communication Checklist (CCC): a method for assessing qualitative aspects of communicative impairment in children. *J. Child Psychol. Psychiatry.* **39,** 879–91 (1998).

29. Scourfield, J., Martin, N., Lewis, G. & McGuffin, P. Heritability of social cognitive skills in children and adolescents. *Br. J. Psychiatry* **175,** (1999).

30. Goodman, R. The Strengths and Difficulties Questionnaire: A Research Note. *J. Child Psychol. Psychiatry* **38,** 581–586 (1997).

31. Levinson, D. F. *et al.* Copy number variants in schizophrenia: confirmation of five previous findings and new evidence for 3q29 microdeletions and VIPR2 duplications. *Am. J. Psychiatry* **168,** 302–16 (2011).

32. Ching, M. S. L. *et al.* Deletions of NRXN1 (neurexin-1) predispose to a wide spectrum of developmental disorders. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* **153B,** 937–47 (2010).

33. Sanders, S. J. *et al.* Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron* **87,** 1215–33 (2015).

34. Schaaf, C. P. *et al.* Phenotypic spectrum and genotype-phenotype correlations of NRXN1 exon deletions. *Eur. J. Hum. Genet.* **20,** 1240–7 (2012).

35. Fromer, M. *et al.* Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat. Neurosci.* **19,** 1442–1453 (2016).

36. Ripke, S. *et al.* Genome-wide association study identifies five new schizophrenia loci.

21

Nat. Genet. **43,** 969–976 (2011).

37.   Genome-Wide Association Study Implicates HLA-C*01:02 as a Risk Factor at the Major Histocompatibility Complex Locus in Schizophrenia. *Biol. Psychiatry* **72,** 620–628 (2012).

38.   Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* **381,** 1371–9 (2013).

39.   Ripke, S. *et al.* Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511,** 421–7 (2014).

40.   Nagel, M. *et al.* GWAS Meta-Analysis of Neuroticism (N=449,484) Identifies Novel Genetic Loci and Pathways. *bioRxiv* (2017).

41.   Okbay, A. *et al.* Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat. Genet.* **48,** 624–633 (2016).

42.   Sniekers, S. *et al.* Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nat. Genet.* (2017). doi:10.1038/ng.3869

43.   Okbay, A. *et al.* Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533,** 539–542 (2016).

44.   Li, N. P. & Kanazawa, S. Country roads, take me home… to my friends: How intelligence, population density, and friendship affect modern happiness. *Br. J. Psychol.* **107,** 675–697 (2016).

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

46.   Kosmicki, J. A. *et al.* Refining the role of de novo protein-truncating variants in

neurodevelopmental disorders by using population reference samples. *Nat. Genet.* **49,** 504–510 (2017).

47. Geurts, H. M. *et al.* Can the Children's Communication Checklist differentiate between children with autism, children with ADHD, and normal controls? *J. Child Psychol. Psychiatry* **45,** 1437–1453 (2004).

48. Weiner, D. J. *et al.* Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. *Nat. Genet.* (2017). doi:10.1038/ng.3863

49. Goodman, R., Ford, T., Simmons, H., Gatward, R. & Meltzer, H. Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *Br. J. Psychiatry* **177,** 534–539 (2000).

**Figure 1: Genetic correlations**
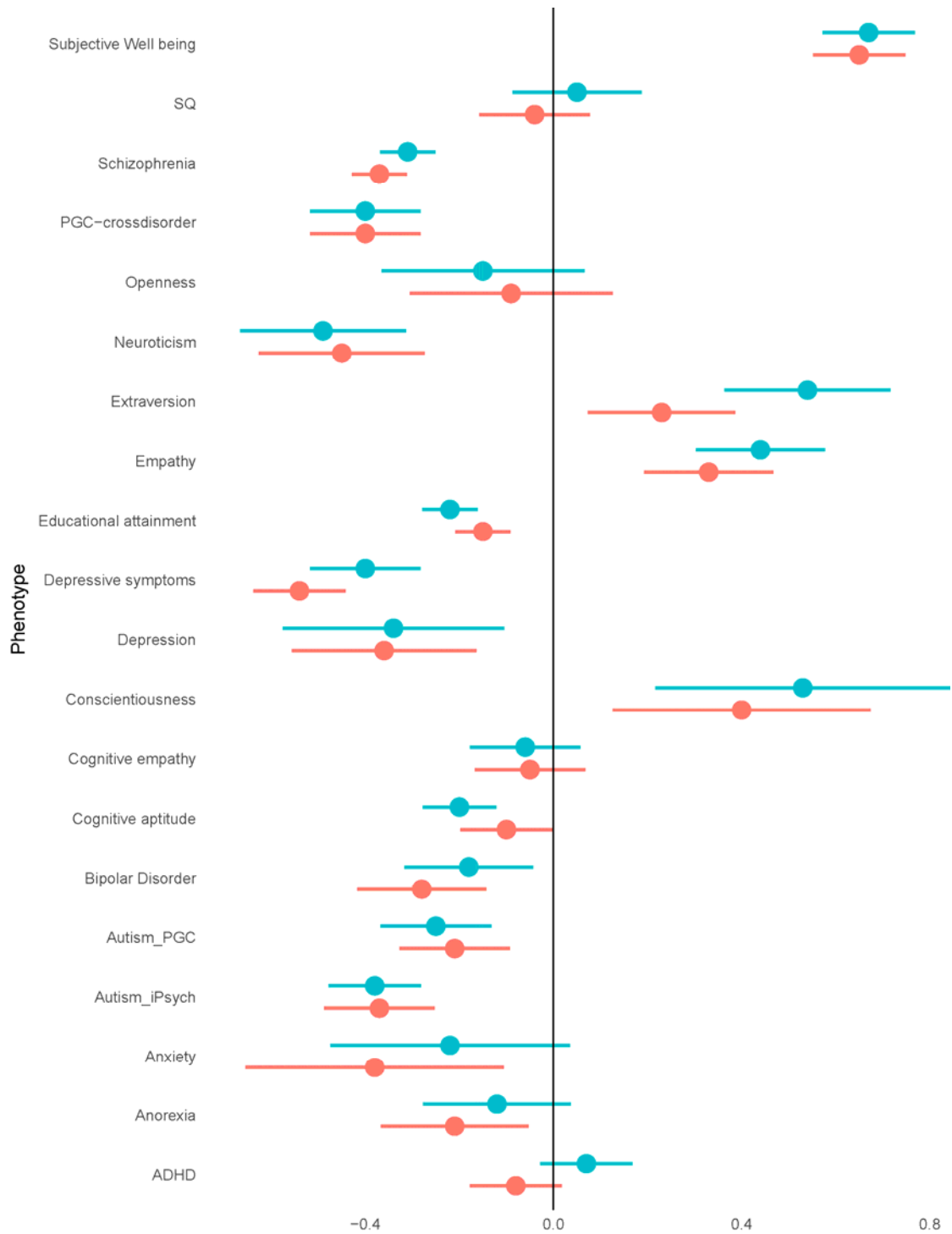
Genetic correlations and 95% confidence intervals provided for the two phenotypes (pink is family relationship satisfaction, blue is friendship satisfaction). Phenotypes tested are on the y axis. Genetic correlation is on the x axis.

Figure 2: Manhattan and QQ-plots
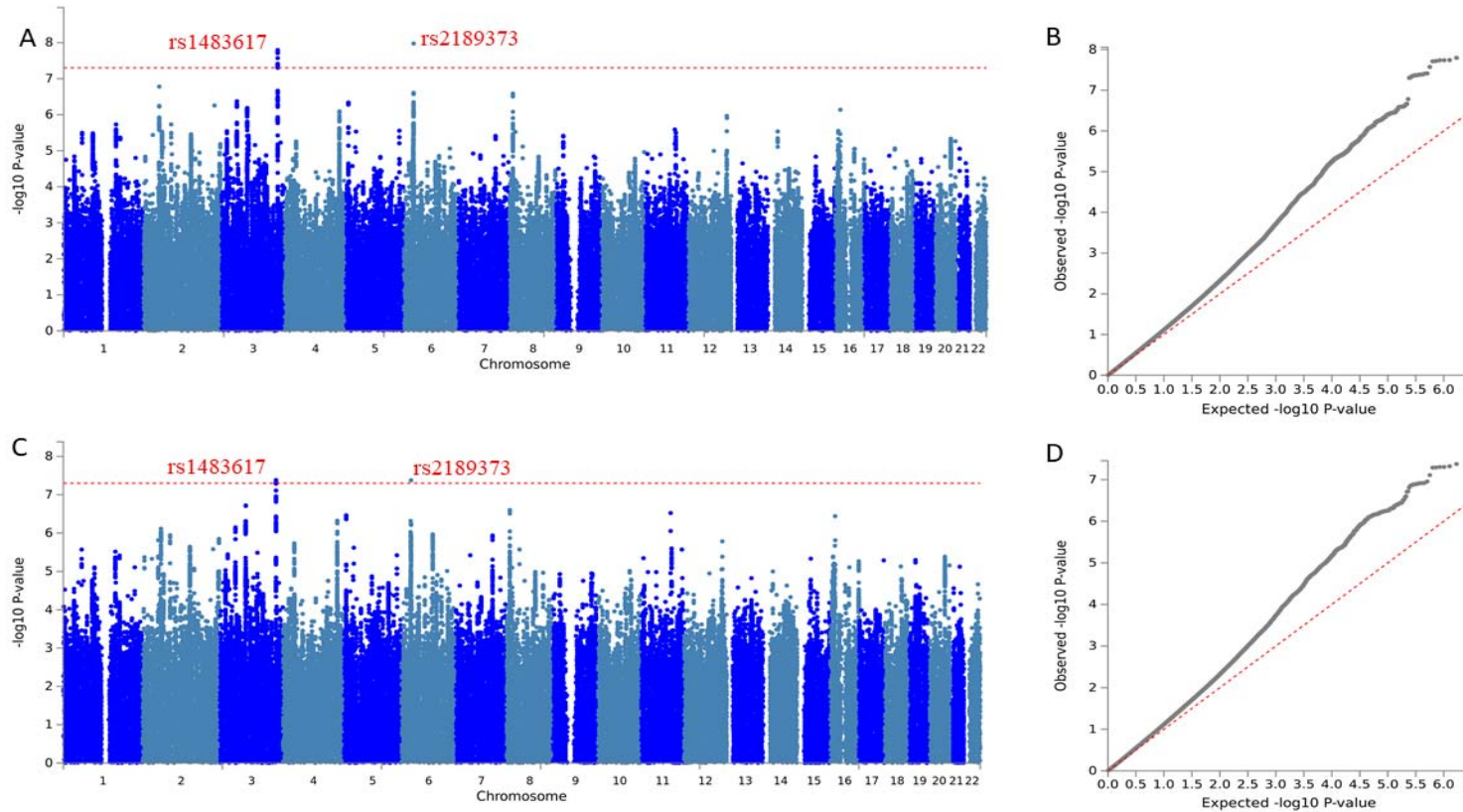
Manhattan and QQplots of the two MTAG-GWAS datasets. A: Manhattan plot of family relationship satisfaction ($N_{effective}$ = 164,112). B: QQplot of family relationship satisfaction (LDSR intercept = 0.99(0.0084); λ_{GC} = 1.18). C. Manhattan plot for friendship satisfaction $N_{effective}$ = 158,116). D QQplot for friendship satisfaction (LDSR intercept = 0.99(0.0087); λ_{GC} = 1.17). Significant SNPs have been labelled on the Manhattan plots in red.

Figure 3: Circos plots showing chromosome interactions and eQTLs
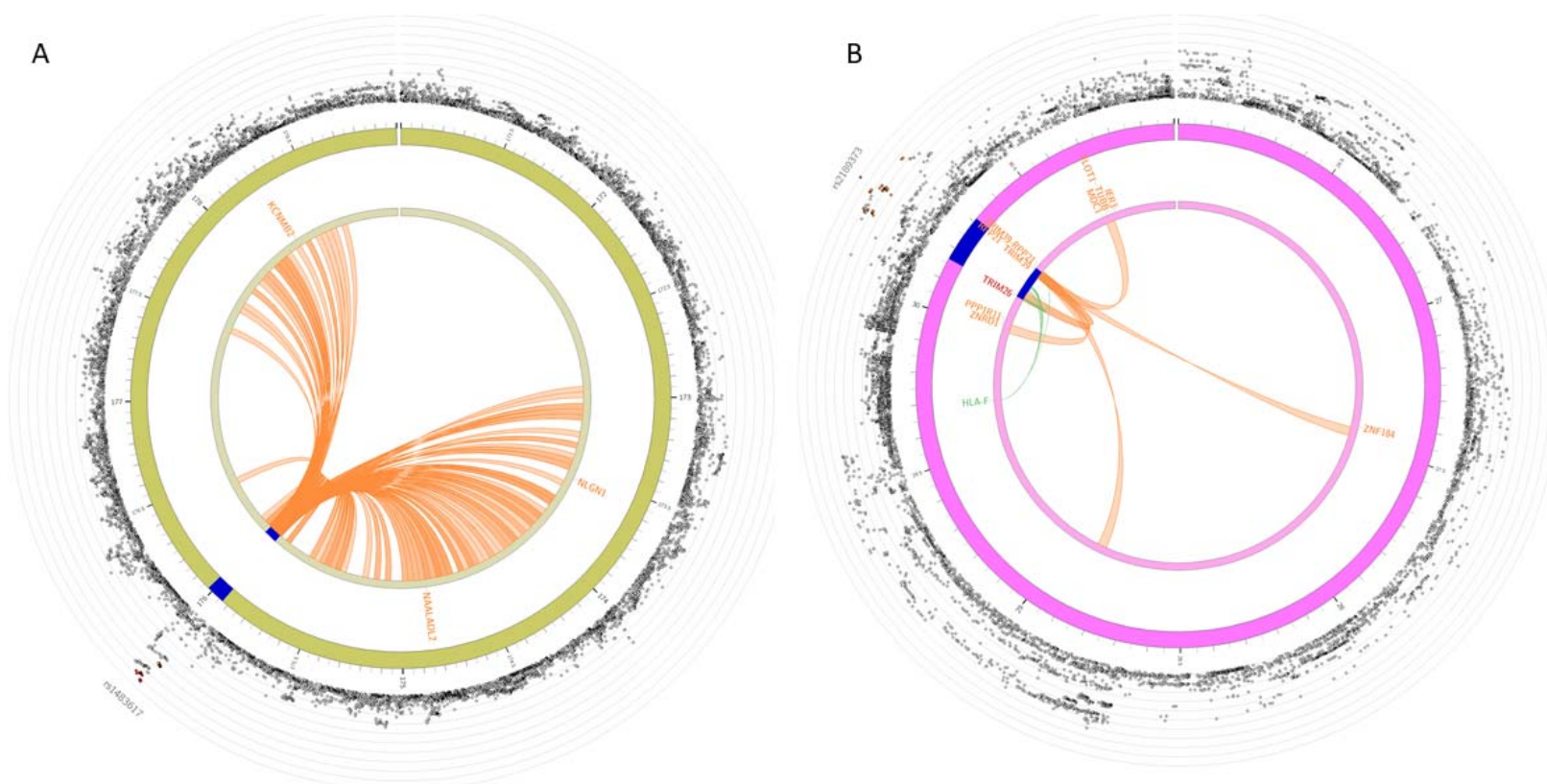
Circos plots showing eQTLs (green lines) and chromosome interactions (orange lines) for the two significant loci. A: Circos plot for rs1483617. B: Circos plot for rs2189373. The outer ring shows a modified Manhattan plot with -log10 P-values. Blue regions identify genomic risk regions. Names of genes with significant chromosome interactions or eQTL interactions are provided in the middle circle. Red regions indicate regions implicated by both eQTL and chromosome interactions.