Identifying loci with different allele frequencies among cases of eight psychiatric disorders using CC-GWAS

Wouter J. Peyrot[1,2,3,*], **Alkes L. Price**[1,2,4,*]

[1]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA. [2]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [3]Department of Psychiatry, Amsterdam UMC, Vrije Universiteit, Amsterdam 1081 HJ, Netherlands. [4]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA

Abstract

Psychiatric disorders are highly genetically correlated, but little research has been conducted on the genetic differences between disorders. We developed a new method (CC-GWAS) to test for differences in allele frequency among cases of two disorders using summary statistics from the respective case-control GWAS, transcending current methods that require individual-level data. Simulations and analytical computations confirm that CC-GWAS is well-powered with effective control of type I error. We applied CC-GWAS to publicly available summary statistics for schizophrenia, bipolar disorder, major depressive disorder, and five other psychiatric disorders. CC-GWAS identified 196 independent case-case loci, including 72 CC-GWAS-specific loci that were not genome-wide significant in the input case-control summary statistics; two of the CC-GWAS-specific loci implicate the genes *KLF6* and *KLF16* (from the Kruppel-like family of transcription factors), which have been linked to neurite outgrowth and axon regeneration. CC-GWAS loci replicated convincingly in applications to data sets with independent replication data.

Introduction

Psychiatric disorders are highly genetically correlated, and many studies have focused on their shared genetic components, including genetic correlation estimates of up to 0.7[1–3] and recent identification of 109 pleotropic loci across a broad set of eight psychiatric disorders[3]. However, much less research has been conducted on the genetic differences between psychiatric disorders, and biological differences between psychiatric disorders are poorly understood. Currently, differential diagnosis between disorders is often challenging and

treatments are often non-disorder-specific, highlighting the importance of studying genetic differences between psychiatric disorders.

A recent study[4] progressed the research on genetic differences between disorders by comparing individual-level data of 24k schizophrenia (SCZ) cases vs. 15k bipolar disorder (BIP) cases, yielding two significantly associated loci. However, ~25% of the cases were discarded compared to the respective case-control data (owing to non-matching ancestry and genotyping platform). Methods that analyse case-control summary statistics may be advantageous, because they make use of all genotyped samples and because summary statistics are often broadly publicly available[5]. Indeed, several methods have been developed to analyse genome-wide association study (GWAS) summary statistics of two complex traits[3,6–14], but none of these methods can be used to conduct a case-case comparison of allele frequencies among cases (see Discussion). Currently, case-case comparisons of allele frequencies among cases of two disorders require individual-level data from cases of both disorders.

In this study, we propose a new method (CC-GWAS) to compare cases of two disorders based on the respective case-control GWAS summary statistics (while modelling sample overlap[6]). CC-GWAS relies on a new genetic distance measure ($F_{ST,causal}$) quantifying the genetic distances between cases and controls of different disorders. We first apply CC-GWAS to publicly available GWAS summary statistics of the mood and psychotic disorders[3], SCZ[15,16], BIP[17] and major depressive disorder (MDD)[18]. Subsequently, we analyse all comparisons of eight psychiatric disorders by additionally analysing attention deficit/hyperactivity disorder (ADHD)[19], anorexia nervosa (AN)[20], autism spectrum disorder (ASD)[21], obsessive–compulsive disorder (OCD)[22], Tourette's Syndrome and Other Tic Disorders (TS)[23]. Finally, we replicate CC-GWAS results in applications to data sets for which independent replication data were available.

Results

Overview of methods

CC-GWAS detects differences in allele frequencies among cases of two different disorders A and B by analysing case-control GWAS summary statistics for each disorder. CC-GWAS relies on the analytical variances and covariances of genetic effects of causal variants distinguishing caseA vs. controlA (A1A0), caseB vs. controlB (B1B0), and caseA vs. caseB (A1B1).

CC-GWAS weights the effect sizes from the respective case-control GWAS using weights that minimize the expected squared difference between estimated and true A1B1 effect sizes; we refer to these as CC-GWAS ordinary least squares (CC-GWAS$_{OLS}$) weights (see Methods). The CC-GWAS$_{OLS}$ weights are designed to optimize power to detect A1B1, and depend on sample size, sample overlap, and the expected variances and covariances of effect sizes. The CC-GWAS$_{OLS}$ weights may be susceptible to type I error for SNPs with nonzero A1A0 and B1B0 effect sizes but zero A1B1 effect size, which we refer to as "stress test" SNPs. To mitigate this, CC-GWAS also computes sample size independent weights based on infinite sample size; we refer to these as CC-GWAS Exact (CC-GWAS$_{Exact}$) weights (see

Methods). CC-GWAS reports a SNP as statistically significant if it achieves $P<5\times10^{-8}$ using CC-GWAS$_{OLS}$ weights *and* $P<10^{-4}$ using CC-GWAS$_{Exact}$ weights: the CC-GWAS$_{OLS}$ weights optimizes power and protect against type I error at null-null SNPs (with no impact on either disorder), while using the CC-GWAS$_{Exact}$ weights protects against type I error at stress test SNPs. For statistically significant SNPs, CC-GWAS outputs CC-GWAS$_{OLS}$ effect sizes reflecting direction and magnitude of effect. Importantly, CC-GWAS identifies and discards false positive associations that can arise due to differential tagging of a causal stress test SNP, e.g. due to subtle differences in ancestry between the input case-control studies (see Methods). When there is substantial uncertainty in population prevalence, a range of possible disorder prevalences can be specified. We further note that sample overlap of controls increases the power of CC-GWAS, by providing a more direct comparison of caseA vs. caseB. Further details of the CC-GWAS method are provided in the Methods section; we have released open-source software implementing the method (see Code availability section).

The CC-GWAS$_{OLS}$ weights depend on a population-level quantity that we refer to as $F_{ST,causal}$, the average normalized squared difference in allele frequencies of causal variants. $F_{ST,causal}$ is derived based on the SNP-based heritabilities ($h^2_{l,A}$ and $h^2_{l,B}$), lifetime population prevalences (K_A and K_B), genetic correlation (r_g), and number of independent causal variants (m). $F_{ST,causal}$ allows for a direct comparison of cases and controls using $\sqrt{m * F_{ST,causal}}$ as a genetic distance measure, where the square root facilitates 2-dimensional visualization (Figure 1 and Methods).

Main simulations

We assessed the power and type I error of CC-GWAS using both simulations with individual-level data and analytical computations (see Methods). We compared four methods: CC-GWAS; the CC-GWAS$_{OLS}$ component of CC-GWAS; the CC-GWAS$_{Exact}$ component of CC-GWAS; and a simple method that uses weight +1 for A1A0 and −1 for B1B0 (Delta method; mentioned in ref.[6]). We assessed (i) power to detect causal SNPs with case-control effect sizes for both disorders drawn from a bivariate normal distribution (allele frequencies A0 A1, B0 B1, A1 B1); (ii) type I error for "null-null" SNPs, defined as SNPs with no effect on either disorder (A0=A1, B0=B1, A1=B1); and (iii) type I error for "stress test" SNPs, defined as SNPs with A0 A1, B0 B1, A1=B1 (see above). Default parameter settings were loosely based on the genetic architectures of SCZ and MDD with liability-scale h^2=0.2, prevalence K=0.01, and sample size 100,000 cases + 100,000 controls for disorder A; liability-scale h^2=0.1, prevalence K=0.15, and sample size 100,000 cases + 100,000 controls for disorder B; genetic correlation r_g=0.5 between disorders; and m=5,000 causal SNPs affecting both disorders with causal effect sizes following a bivariate normal distribution. For these parameter settings, the weights are 0.86 for A1A0 and −0.55 for B1B0 for the CC-GWAS$_{OLS}$ component, and 0.99 and −0.85 respectively for the CC-GWAS$_{Exact}$ component. The CC-GWAS$_{OLS}$ component assigns relatively more weight to A1A0 (0.86/0.55=1.56) than the CC-GWAS$_{Exact}$ component (0.99/0.85=1.16), because of the larger heritability and lower prevalence of A1A0 (implying higher signal to noise ratio at the same case-control sample size[24,25]). Each stress test SNP was specified to explain 0.10% of liability-scale variance in A and 0.29% of liability-scale variance in B (resulting in allele

frequency A1=B1); we focused on large-effect stress test SNPs to provide a maximally stringent assessment of the robustness of CC-GWAS to stress test SNPs. Other parameter settings were also explored.

Results of analytical computations are reported in Figure 2 and Supplementary Table 1; simulations with individual-level data produced identical results (Supplementary Table 2), thus we focus primarily on results of analytical computations. We reached three main conclusions. First, CC-GWAS attains similar power as the CC-GWAS$_{OLS}$ component by itself, higher power than the CC-GWAS$_{Exact}$ component by itself, and much higher power than the Delta method (Figure 2A); we note that this is a best-case scenario for CC-GWAS, as the simulated bivariate genetic architecture follows the CC-GWAS assumptions. As expected, power increases with increasing sample size and decreases with increasing genetic correlation. The power of CC-GWAS to detect case-case differences lies in between the power of the input A1A0 and B1B0 summary statistics to detect case-control differences (Supplementary Figure 1). Second, all methods perfectly control type I error at null-null SNPs, with a per-SNP type I error rate $< 5 \times 10^{-8}$ (Figure 2B). Third, although the CC-GWAS$_{OLS}$ component has a severe type I error problem at stress test SNPs (particularly when the genetic correlation is large), CC-GWAS (incorporating the CC-GWAS$_{Exact}$ component) attains effective control of type I error at stress test SNPs (per-SNP type I error rate $< 10^{-4}$; Figure 2C), an extreme category of SNPs that is not likely to occur often in empirical data. Notably, with increasing sample size the CC-GWAS$_{OLS}$ weights converge towards the CC-GWAS$_{Exact}$ weights (Supplementary Figure 2). In conclusion, CC-GWAS balances the high power of the CC-GWAS$_{OLS}$ component with effective control of type I error of the CC-GWAS$_{Exact}$ component. We discuss secondary simulation results in the Supplementary Note, Supplementary Tables 3–5 and Supplementary Figures 3–12.

Assessing the robustness of CC-GWAS

We performed two sets of analyses to further assess the robustness of CC-GWAS. First, we assessed the robustness of CC-GWAS to false positive associations due to differential tagging of a causal stress test SNP (Figure 3A). CC-GWAS screens the region around every genome-wide significant candidate CC-GWAS SNP for evidence of a differentially linked stress test SNP, and conservatively filters the candidate CC-GWAS SNP when suggestive evidence of a differentially linked stress test SNP is detected (see Methods and Supplementary Table 6). We simulated GWAS results of loci with a causal stress test SNP using real LD patterns in two distinct European populations (see Methods). We compared four methods/scenarios: CC-GWAS (with filtering) in the scenario where the causal stress test SNP is genotyped/imputed; CC-GWAS (with filtering) in the scenario where the causal stress test SNP is *not* genotyped/imputed; CC-GWAS with no filtering (for which it is irrelevant whether the causal stress test SNP is genotyped/imputed); and direct case-case GWAS (with no filtering). We report the per-locus type I error rate: the number of loci with at least one genome-wide significant tagging SNP divided by the number of loci tested. The parameter settings were identical to those of the stress test SNPs in Figure 2C.

Results are reported in Figure 3B and Supplementary Table 7. We reached four main conclusions. First, CC-GWAS (with filtering) attained effective control of type I error, with

per-locus type I error rate $<5\times10^{-8}$ in the scenario where the causal stress test SNP is genotyped/imputed. Second, CC-GWAS (with filtering) also attained effective control of type I error, with per-locus type I error rate $<10^{-4}$, in the scenario where the causal stress test SNP is *not* genotyped/imputed. Analogous to our main simulations above, we note that stress test SNPS form an extreme category of SNPs that is not likely to occur often in empirical data. Third, CC-GWAS with no filtering suffered a per-locus type I error rate of up to 0.07, underscoring the necessity of applying the filter in CC-GWAS. Fourth, the direct case-case GWAS suffered a per-locus type I error rate of up to 0.09. Thus, the robustness of CC-GWAS (with filtering) to differential tagging represents a major improvement over direct case-case GWAS (with no filtering). We discuss secondary simulation results in the Supplementary Note and Supplementary Tables 7–8.

Second, we applied CC-GWAS to two sets of empirical case-control GWAS summary statistics for the same disorder, for which no case-case associations are expected. We focused on breast cancer (BC), as this is a disorder with two sets of independent, publicly available GWAS summary statistics in very large, well-powered samples[26] (see Methods). The CC-GWAS analyses yielded no genome-wide significant case-case association (Supplementary Table 9). Notably, CC-GWAS identified two genome-wide significant candidate loci prior to filtering for differential tagging of stress test SNPs. All SNPs in these two loci very clearly met the filtering criteria (Supplementary Table 10), and remained filtered when applying various perturbations to the filtering criteria (Supplementary Table 8). We emphasize that these BC vs. BC analyses were only intended to test the robustness of CC-GWAS, as CC-GWAS is intended for comparing two *different* disorders with genetic correlation < 0.8 (see Methods). Nevertheless, our analyses of BC vs. BC further validate the robustness of CC-GWAS.

CC-GWAS identifies 116 independent loci with different allele frequencies among cases of SCZ, BIP and MDD

We applied CC-GWAS to publicly available summary statistics for SCZ,[16] BIP[17] and MDD[18] (Table 1). To run CC-GWAS, we assumed 10,000 independent causal SNPs for each psychiatric disorder[27] (see Methods for a detailed discussion of the assumed number of causal SNPs in applications of CC-GWAS). The underlying CC-GWAS$_{OLS}$ weights and CC-GWAS$_{Exact}$ weights used by CC-GWAS are reported in Table 1, along with the disorder-specific parameters used to derive these weights. The CC-GWAS$_{OLS}$ weights are based on the expected genetic distances between cases and/or controls ($F_{ST,causal}$) (Figure 1B–D and Supplementary Table 11). For each disorder, we specified a range of disorder prevalences to CC-GWAS$_{Exact}$ (Table 1; see Overview of Methods). We defined *CC-GWAS-specific* loci as loci for which none of the genome-wide significant SNPs had $r^2>0.8$ with any of the genome-wide significant SNPs in the input case-control GWAS results (Supplementary Table 12).

For each pair of SCZ, BIP and MDD, the total number of independent CC-GWAS loci and number of independent CC-GWAS-specific loci are reported in Table 1. The CC-GWAS analysis identified 121 loci summed across pairs of disorders, resulting in 116 independent CC-GWAS loci (Supplementary Table 12) including 21 CC-GWAS-specific loci; 8 of these

loci have not been reported previously (conservatively defined as: no SNP in 1000 Genomes[28] with $r^2>0.8$ with a genome-wide significant CC-GWAS SNP in the locus reported for any trait in the NHGRI GWAS Catalog[29]; Supplementary Table 12). Notably, the CC-GWAS$_{Exact}$ component did not filter out any variants identified using the CC-GWAS$_{OLS}$ component, i.e. all SNPs with $P<5\times10^{-8}$ using CC-GWAS$_{OLS}$ weights also had $P<10^{-4}$ using CC-GWAS$_{Exact}$ weights (for all disorder prevalences in the specified ranges), because the CC-GWAS$_{OLS}$ weights were relatively balanced. In addition, no variants were excluded based on the filtering step to exclude potential false positive associations due to differential tagging of a causal stress test SNP. For each CC-GWAS locus, the respective input case-control effect sizes for each disorder are reported in Figure 4 and Supplementary Table 13. Details of the 21 CC-GWAS-specific loci are reported in Table 2, and details of the remaining 100 CC-GWAS loci are reported in Supplementary Table 13 (the locus names reported in these Tables incorporate results of our SMR analysis[30]; see below). We discuss secondary analyses in the Supplementary Note, Supplementary Tables 14–22 and Supplementary Figure 11.

CC-GWAS-specific loci implicate known and novel disorder genes

We used two approaches to link the 21 CC-GWAS-specific loci to genes (Table 2). First, we linked exonic lead SNPs to the corresponding genes. Second, we used the SMR test for colocalization[30] to identify CC-GWAS loci with significant associations between gene expression effect sizes in *cis* across 14 brain tissues[31,32] and CC-GWAS$_{OLS}$ case-case effect sizes (see Methods and Supplementary Table 23). Below, we highlight 4 CC-GWAS-specific loci from Table 2, representing both known and novel findings.

The CC-GWAS-specific SCZ vs. MDD locus defined by lead SNP rs2563297 (chr5:140,097,072) produced significant SMR colocalization results for 11 gene-tissue pairs representing 7 unique genes (Supplementary Table 23). The 7 unique genes included 5 protocadherin alpha (*PCDHA*) genes, which play a critical role in the establishment and function of specific cell-cell connections in the brain[33], and the *NDUFA2* gene, which has been associated with Leigh syndrome (an early-onset progressive neurodegenerative disorder)[34]. Significant CC-GWAS SNPs in this locus have previously been associated to schizophrenia[35–38] (in data sets distinct from our input schizophrenia GWAS[16], in which this locus was not significant due to sampling variance and/or ancestry differences), depressive symptoms[7], neuroticism[39], educational attainment[38,40], intelligence[41], blood pressure[42,43], and a meta-analyses of schizophrenia, education and cognition[38], implying that this is a highly pleiotropic locus.

The CC-GWAS-specific SCZ vs. MDD locus defined by lead SNP rs2944833 (chr7:71,774,496) produced a significant SMR colocalization result for one gene-tissue pair, involving the *CALN1* gene in meta-analyzed brain eQTL[32] (Supplementary Table 23). *CALN1* plays a role in the physiology of neurons and is potentially important in memory and learning[44]. Indeed, SNPs in this locus have previously been associated to educational attainment[40,45], intelligence[41,46], cognitive function[47], and a meta-analysis of schizophrenia, education and cognition[38]. Again, this implies that CC-GWAS can increase power to detect associated loci in the input case-control GWAS data sets analyzed here.

Finally, two distinct CC-GWAS-specific loci implicated genes in the Kruppel-like family of transcription factors. The CC-GWAS-specific SCZ vs. BIP locus defined by lead SNP rs1054972 (chr19:1,852,582) located within an exon of *KLF16*, and the CC-GWAS-specific SCZ vs. MDD locus defined by lead SNP rs17731 (chr10:3,821,561) located within an exon of *KLF6*. The respective case-control effect sizes suggest that rs1054972 and rs17731 both have an impact on SCZ, but have not yet reached significance in the respective case-control analyses (P=1.3e–5 and P=2.9e–7 respectively; Table 2 and Supplementary Table 24). *KLF16* and *KLF6* play a role in DNA-binding transcription factor activity and in neurite outgrowth and axon regeneration[48], and we hypothesize they may play a role in the previously described schizophrenia pathomechanism of synaptic pruning[49]. Furthermore, the *KLF5* gene from the same gene family has previously been reported to be downregulated in post-mortem brains of schizophrenia patients[50]. At the time of our analyses, *KLF16* and *KLF6* had not previously been associated to schizophrenia; *KLF6* has very recently been associated to schizophrenia in a meta-analysis of East Asian and European populations[37], but *KLF16* has still not been associated to schizophrenia. This implies that CC-GWAS can identify novel disorder genes.

CC-GWAS identifies 196 independent loci distinguishing cases of eight psychiatric disorders

We applied CC-GWAS to all 28 pairs of eight psychiatric disorders by analysing ADHD[19], AN[20], ASD[21], OCD[22], and TS[23] in addition to SCZ[16], BIP[17] and MDD[18] (Supplementary Table 25). To run CC-CWAS, we assumed 10,000 independent causal SNPs for each psychiatric disorder[27] (see Methods). The underlying CC-GWAS$_{OLS}$ weights and CC-GWAS$_{Exact}$ weights used by CC-GWAS are reported in Supplementary Table 14. The CC-GWAS$_{OLS}$ weights are based on the expected genetic distances between cases and/or controls ($F_{ST,causal}$) (Supplementary Figure 13 and Supplementary Table 11). For each disorder, we specified a range of disorder prevalences to CC-GWAS$_{Exact}$ (Supplementary Table 25). For each pair of psychiatric disorders, the total number of independent CC-GWAS loci and number of independent CC-GWAS-specific loci are reported in Table 3. The CC-GWAS analysis identified 313 loci summed across pairs of disorders, resulting in 196 independent loci including 72 CC-GWAS-specific loci. Seven candidate CC-GWAS loci were excluded (reducing the number of CC-GWAS loci from 320 to 313) based on the filter to exclude potential false positive associations due to differential tagging of a causal stress test SNP (Supplementary Note, Supplementary Tables 6 and 10). A further detailed description of results is provided in the Supplementary Note, Supplementary Tables 8,12–14,23,25 and Supplementary Figure 14.

CC-GWAS loci replicate in independent data sets

We investigated whether case-case associations identified by CC-GWAS replicate in independent data sets. Of the eight psychiatric disorders, only SCZ and MDD had sufficient sample size to perform replication analyses of the SCZ vs. MDD results based on publicly available GWAS results of subsets of the data[15,51]. To further validate the CC-GWAS method, we also analysed three case-case comparisons of three autoimmune disorders with publicly available GWAS results for independent discovery and replication data sets with substantial sample sizes (Crohn's disorder (CD)[52], ulcerative colitis (UC)[52] and rheumatoid

arthritis (RA)[53]. To obtain independent replication results for SCZ, MDD, CD and UC, we subtracted the subset GWAS results from the full set GWAS results[16,18,52] using MetaSubtract[54].

Results for these four pairs of disorders are reported in Figure 5, Supplementary Tables 26–27. For SCZ vs. MDD, the CC-GWAS discovery analysis identified 57 independent loci (less than the 99 independent loci in Table 1, due to smaller sample size), yielding a replication slope (based on a regression of replication vs. discovery effect sizes[55]) of 0.57 (SE 0.06) (Figure 5A) comparable to the corresponding case-control replication slopes (Supplementary Figure 15, and Supplementary Tables 26,28). We hypothesize that all slopes were smaller than 1 owing to within-disorder heterogeneity[1]. For the autoimmune disorders, a replication slope of 0.83 (SE 0.03) was obtained (Figure 5B), again comparable to the corresponding case-control replication slopes (Supplementary Figure 15). We further investigated the replication of the subset of 22 CC-GWAS-specific loci (9 for SCZ vs. MDD and 13 for the 3 autoimmune disorders) obtaining a replication slope of 0.70 (SE 0.07) (Figure 5C), which was borderline significantly different (P=0.07) from the slope of 0.83 (0.02) for the 97 remaining loci (Figure 5D); we note that CC-GWAS-specific loci had smaller case-case effect sizes and are thus expected to be more susceptible to winner's curse[56] and attain a lower replication slope.

We performed additional replication analyses that did not require the use of MetaSubtract, analysing 6 comparisons of each pair of 4 disorders (of low biological interest, but useful for assessing the robustness of the CC-GWAS method): the resulting 153 CC-GWAS loci replicated convincingly (Supplementary Figures 16). A further detailed description of the replication analyses is provided in the Supplementary Note, Supplementary Tables 25–28, and Supplementary Figures 13, 15–16.

Discussion

We developed a new method, CC-GWAS, to compare allele frequencies among cases of two different disorders by analysing case-control GWAS summary statistics for each disorder (while modelling sample overlap[6]). We identified 196 independent loci with different allele frequencies among cases of eight psychiatric disorders by applying our CC-GWAS method to the respective case-control GWAS summary statistics. 72 of the 196 loci were CC-GWAS-specific, highlighting the potential of CC-GWAS to produce new biological insights. In particular, the lead SNPs of two distinct loci were located in exons of *KLF6* and *KLF16*, which have been linked to neurite outgrowth and axon regeneration[48]; we hypothesize that these genes may be involved in the role of synaptic pruning in SCZ[49]. We confirmed the robustness of CC-GWAS via simulations, analytical computations, empirical analysis of BC vs. BC, and independent replication of empirical CC-GWAS results.

A detailed discussion of differences between CC-GWAS and other methods based on summary statistics – the Delta method[6]; methods combining GWAS results of correlated disorders to increase power[7–9]; GWIS[10]; ASSET[11]; and *mtCOJO*[13] – is provided in the Supplementary Note, Supplementary Tables 4,13, and Supplementary Figures 17 and 18. The most natural method to compare CC-GWAS to is a case-case GWAS based on

individual-level data, as performed in ref.[4] for SCZ vs. BIP based on individual level data from ref.[15] and ref.[17] respectively. CC-GWAS identified 12 SCZ vs. BIP loci (or 10 when applied to data from ref.[15] and ref.[17], as in ref.[4]) compared to 2 SCZ vs. BIP loci identified in ref.[4], which discarded ~25% of the cases compared to the respective case-control data (owing to non-matching ancestry and genotyping platform). The results of ref.[4] and CC-GWAS were generally concordant (see Supplementary Note and Supplementary Tables 28,29). We note two advantages of CC-GWAS over a direct case-case GWAS. First, CC-GWAS is much less sensitive to subtle allele frequency differences due to differences in ancestry and/or genotyping platform (Supplementary Table 30), because the case-case comparison accounts for the allele frequency in matched controls by comparing case-control effects. Second, CC-GWAS filters potential false positive associations due to differential tagging of a causal stress test SNP (with the same allele frequency in cases of both disorders). This is not possible in a direct case-case GWAS based on data from cases alone, as the filtering criteria require information about case-control effect sizes.

The CC-GWAS method has several limitations. First, the choice of the threshold for the CC-GWAS$_{Exact}$ p-values in CC-GWAS is somewhat arbitrary, but we believe 10^{-4} is a reasonable choice as it (i) effectively protects against false positives due to stress test SNPs (Figure 2C and Supplementary Figure 6), which cannot be numerous (e.g. 100 independent stress test SNPs as defined in Figure 2C would explain 29% of liability-scale variance in disorder B), and (ii) has only limited impact on the power of CC-GWAS (Figure 2A). Second, the filtering criteria to avoid false positive associations due to differential tagging of a causal stress test SNP (Supplementary Table 6) are ad hoc and somewhat arbitrary. However, we verified that applying perturbations to these filtering criteria had little impact on our results, both in extensive simulations (Supplementary Table 8) and in analyses of empirical data (Supplementary Table 8). We discuss ten additional limitations in detail in the Supplementary Note, Supplementary Tables 5–8,14,31–34, and Supplementary Figures 6 and 19.

In conclusion, we have shown that CC-GWAS can reliably identify loci with different allele frequencies among cases, providing novel biological insights into the genetic differences between cases of eight psychiatric disorders. Thus, CC-GWAS helps promote the ambitious but important goal of better clinical diagnoses and more disorder-specific treatment of psychiatric disorders.

Methods

Quantifying genetic distances between cases and/or controls of each disorder

We derive $F_{ST,causal}$, the average normalized squared difference in allele frequencies at independent causal variants, for the comparisons $A1A0$, $B1B0$, $A1B1$, $A1B0$, $A0B1$, $A0B0$. We consider two disorders A and B with lifetime prevalences K_A and K_B, liability-scale heritabilities h^2_{lA} and h^2_{lB}, and genetic correlation r_g. Assume the heritabilities and genetic correlation have been assessed on data of m independent SNPs, and assume these SNPs impact both traits with effects following a bivariate normal distribution. The heritabilities are transposed to the observed scales, h^2_{oA} and h^2_{oB}, with proportions of cases of 0.5 in line with

refs.[25,57]. The coheritability is also expressed on this scale as $coh_{oA, oB} = r_g\sqrt{h_{oA}^2\, h_{oB}^2}$. We derive $F_{ST,causal,A1A0}$ as (see Supplementary Note and Supplementary Tables 35,36)

$$F_{ST,\, causal,\, A1A0} \approx \frac{h_{oA}^2}{m}$$

$$F_{ST,\, causal,\, B1B0} \approx \frac{h_{oB}^2}{m}$$

$$F_{ST,\, causal,\, A1B1} \approx \frac{h_{oA}^2}{m}(1 - K_A)^2 - 2\frac{coh_{oA,\, oB}}{m}(1 - K_A)(1 - K_B) + \frac{h_{oB}^2}{m}(1 - K_B)^2$$

$$F_{ST,\, causal,\, A1B0} \approx \frac{h_{oA}^2}{m}(1 - K_A)^2 + 2\frac{coh_{oA,\, oB}}{m}(1 - K_A)K_B + \frac{h_{oB}^2}{m}K_B^2$$

$$F_{ST,\, causal,\, A0B1} \approx \frac{h_{oA}^2}{m}K_A^2 + 2\frac{coh_{oA,\, oB}}{m}K_A(1 - K_B) + \frac{h_{oB}^2}{m}(1 - K_B)^2$$

$$F_{ST,\, causal,\, A0B0} \approx \frac{h_{oA}^2}{m}K_A^2 - 2\frac{coh_{oA,\, oB}}{m}K_A K_B + \frac{h_{oB}^2}{m}K_B^2$$

The quantity $\sqrt{m * F_{ST,\, causal}}$ can be used to represent the cases and controls of two disorders in a 2-dimensional plot (see Figure 1, Supplementary Note and Supplementary Figure 13). We note that the distances can be intuitively interpreted as (i) the square root of the average squared difference in allele frequency at causal SNPs, (ii) proportional to the average power in GWAS (assuming equal sample sizes and numbers of causal SNPs), (iii) heritability on the observed scale based on 50/50 ascertainment (although the heritability has no clear interpretation when comparing overlapping small subsets of the population), and (iv) an indication of the accuracy of polygenic risk prediction. The genetic correlation r_g is equal to the cosines of the angle of the lines (population mean - $A1$) and (population mean - $B1$).

In application, we derive $F_{ST,causal}$ analytically based on the heritabilities, population prevalences and genetic correlation. We note three important differences between $F_{ST,causal}$ and the F_{ST} from population genetics[58]. First, we restrict our definition of $F_{ST,causal}$ to independent SNPs, while F_{ST} from population genetics is based on all genome-wide SNPs. Second, $F_{ST,causal}$ at variants with large LD-scores are larger than at SNPs with low LD-scores due to tagging. In contrast, the F_{ST} from population genetics is mainly attributable to drift and more or less evenly distributed over the genome (except for small effects of selection). Third, $F_{ST,causal}$ between cases and controls is of the order of magnitude of 10^{-6}

depending on the number of SNPs m considered. In contrast, the F_{ST} between European and East Asian has been estimated[58] at 0.11. Because of the low magnitude of $F_{ST,causal}$, we report $m * F_{ST,causal}$ in Figure 1 and Supplementary Figure 13 (note that $m * F_{ST,causal}$ is independent of m when other parameters are fixed, because the equations for $F_{ST,causal}$ have m in the denominator; see Supplementary Note).

CC-GWAS method

The CC-GWAS method relies on $F_{ST,causal}$, and assumes that all m SNPs impact both disorders with effect sizes following a bivariate normal distribution (violation of this assumption may impact power, but does not affect type I error rate; see further). CC-GWAS weights the effect sizes from the respective case-control GWAS using weights that minimize the expected squared difference between estimated and true A1B1 effect sizes (while modelling sample overlap[6]); we refer to these as ordinary least squares (CC-GWAS$_{OLS}$) weights. To obtain the CC-GWAS$_{OLS}$ weights, we analytically derive the expected coefficients of regressing the causal effect sizes $A1B1$ on the GWAS results of $A1A0$ and $B1B1$ (see Supplementary Note)

$$\hat{\beta}^{OLS}_{A1B1} = \omega^{OLS}_{A1A0}\hat{\beta}^{GWAS}_{A1A0} + \omega^{OLS}_{B1B0}\hat{\beta}^{GWAS}_{B1B0}$$

The CC-GWAS$_{OLS}$ weights depend on the number of independent causal SNPs, the heritabilities, population prevalences, the genetic correlation, and the variance and covariance of error terms of the betas (depending on sample sizes N_{A1}, N_{A0}, N_{B1}, N_{B0} and the sample overlap between $A0$ and $B0$).

The CC-GWAS$_{OLS}$ weights may be susceptible to type I error for SNPs with nonzero A1A0 and B1B0 effect sizes but zero A1B1 effect size, which we refer to as "stress test" SNPs (see further). To mitigate this, CC-GWAS also computes sample size independent weights based on infinite sample size; we refer to these as CC-GWAS$_{Exact}$ weights. The CC-GWAS$_{Exact}$ weights depend only on the population prevalences K_A and K_B (see Supplementary Note and Supplementary Table 35)

$$\hat{\beta}^{Exact}_{A1B1} = (1 - K_A)\hat{\beta}^{GWAS}_{A1A0} - (1 - K_B)\hat{\beta}^{GWAS}_{B1B0}$$

The z-values and p-values of the CC-GWAS$_{OLS}$ component and CC-GWAS$_{Exact}$ component are estimated by dividing the beta estimates by their standard errors (see Supplementary Note). CC-GWAS reports a SNP as statistically significant if it achieves $P < 5 * 10^{-8}$ using CC-GWAS$_{OLS}$ weights *and* $P < 10^{-4}$ using CC-GWAS$_{Exact}$ weights (all statistical tests in this paper are two-sided), balancing power and type I error. We note that CC-GWAS is intended for comparing two *different* disorders with genetic correlation < 0.8. At larger genetic correlation, the anticipated number of stress test loci may increase posing a risk of per-study type I error > 0.05, and the CC-GWAS$_{OLS}$ weights may become meaningless when the expected genetic distance between cases is close to 0. We note that $\omega_{A1A0} > 0$ and $\omega_{B1B0} < 0$, which indicates that sample overlap of controls (introducing positive covariance between case-control error terms) will reduce the standard error and increase power of CC-

GWAS. When GWAS results are available for a direct case-case GWAS, $\hat{\beta}_{A1B1}^{GWAS}$, CC-GWAS can be extended to CC-GWAS+ (see Supplementary Note). CC-GWAS transposes the case-control odd ratios from logistic regression to the standardized observed scale based on 50/50 case-control ascertainment[59] for convenience in analytical computations (see Supplementary Note).

Filtering criteria to exclude potential false positive associations due to differential tagging of a causal stress test SNP

CC-GWAS identifies and discards false positive associations that can arise due to differential tagging of a causal stress test SNP (Figure 3A). Specifically, CC-GWAS screens the 1MB region around every genome-wide significant candidate CC-GWAS SNP for evidence of a differentially linked stress test SNP, and conservatively filters the candidate CC-GWAS SNP when suggestive evidence of a differentially linked causal stress test SNP is detected. The criteria of the filtering step were motivated by extensive simulations (Supplementary Table 7). For each candidate CC-GWAS SNP, filtering comprises of three sets of criteria (A, B, and C), and the SNP is discarded when at least one of the three sets of criteria is met. The criteria (A) are intended for intermediate sample sizes, the criteria (B) for relatively small sample sizes, and the criteria (C) for very large sample sizes. See Supplementary Note and Supplementary Tables 6,7 for details.

Main simulations and analytical computations to assess power and type I error of CC-GWAS

We first simulated individual-level data of independent SNPs without LD. We simulated (i) causal SNPs with effect sizes following a bivariate normal distribution to assess power, (ii) null-null SNPs with no effect on both disorders to asses type I error, and (iii) stress test SNPs, impacting both disorders but with no case-case allele frequency difference, to assess type I error (see Supplementary Note). The parameters of simulation were largely in line with those used in Figure 2, but in order to reduce computational time, sample sizes were reduced to $N_{A1} = N_{A0} = N_{B1} = N_{B0} = 4,000$, number of causal SNPs to $m = 1,000$, and required levels of significance were reduced to $p < 0.01$ for the CC-GWAS$_{OLS}$ component and $p < 0.05$ for the CC-GWAS$_{Exact}$ component. Simulation results are displayed in Supplementary Table 2 and match analytical computations (see Supplementary Note). The concordance of simulation and analytical computations confirms that increasing sample sizes and decreasing p-value thresholds in analytical computations in Figure 2 is justified. We also simulated data with a different bivariate architecture, with the distribution of SNP effects in line with the general distribution applied in Frei et al.[60]: 1/3 of causal SNPs have an impact on disorder A only, 1/3 of SNPs have an impact on disorder B only, and 1/3 of SNPs have an impact both disorder A and disorder B (Supplementary Table 5).

Simulation of false positive associations due to differential tagging of a causal stress test SNP

We also simulated GWAS results of causal stress test SNPs and their tagging SNPs to study the impact of potential false positive associations due to differential tagging of a causal stress test SNP. We used real LD patterns in two distinct populations[61]: 25k British UK

Biobank samples and 25k "other European" UK Biobank samples (defined as non-British and non-Irish); we note that the F_{ST} between these two populations is 0.0006[58], which is greater than the range of F_{ST} values in the 3 CC-GWAS comparisons of psychiatric disorders for which in-sample allele frequencies were available to estimate F_{ST} (0.0001–0.0005; Supplementary Table 37). An overview of LD (signed correlation) differences between 25k British UK Biobank samples and 25k "other European" UK Biobank samples is provided in Supplementary Table 38. Other parameters were based on our main stress test SNP simulations (Figure 2C). We refer to the Supplementary Note for further details. The advantage of simulating GWAS summary statistics is that this allows increasing the number of simulation runs and sample size dramatically compared to simulations based on individual-level data.

Based on the simulated GWAS results, we first applied CC-GWAS twice including the filtering step: once with the causal stress test SNP included in the GWAS results (CC-GWAS-causal-typed), and once excluding the causal stress test SNP (and all SNPs in perfect LD in both populations) from the GWAS results (CC-GWAS-causal-untyped). Subsequently, we applied CC-GWAS without the filtering step (CC-GWAS-nofilter), and reported the results from $\hat{\beta}_{i, A1B1}$ (Direct case-case GWAS). We report the per-locus type I error rate: the number of loci with at least one genome-wide significant tagging SNP divided by the number of loci tested. The simulation results are reported in Figure 3B and Supplementary Table 7 We performed several secondary simulation analyses (Supplementary Note and Supplementary Tables 7,8).

Empirical data sets

We compared cases from SCZ[16], BIP[17], MDD[18], ADHD[19], AN[20], ASD[21], OCD[22], and TS[23] based on publicly available case-control GWAS results. To further validate CC-GWAS we also compare cases from BC[26], CD[52], UC[52] and RA[53]. Numbers of cases and controls are listed in Table 1 and Supplementary Table 14. In quality control SNPs were removed with $MAF < 0.01$, $INFO < 0.6$, $N_{eff} < 0.67 * max(N_{eff})$, duplicate SNP names, strand-ambiguous SNPs, and the MHC region (chr6:25,000,000–34,000,000) was excluded due to its compiled LD structure. The transformation of odds ratios to the standardize betas on the observed scale requires N_{eff} (Supplementary Note). For some of the disorders (BIP, MDD, AN and RA), N_{eff} was provided on a SNP-by-SNP basis in publicly available GWAS results. For other disorders (SCZ, ADHD, ASD, OCD, TS, BC, CD and UC), we approximated a genome-wide fixed N_{eff} by summing the N_{eff} of the contributing cohorts as $\sum_{cohorts} \frac{4}{1/N_{case, cohort\ i} + 1/N_{control, cohort\ i}}$. (In principle, applying a fixed N_{eff} for cohorts without SNP-by-SNP N_{eff} information could lead to inaccurate transformation of beta for some SNPs. Therefore, we reran CC-GWAS analyses for SCZ, BIP and MDD with fixed N_{eff} yielding nearly identical results to the primary analyses (with fixed N_{eff} for SCZ and SNP-by-SNP N_{eff} for BIP and MDD). This confirms that using fixed N_{eff} is appropriate.) All reported SNP names and chromosome positions are based on GRCh37/hg19.

Application of CC-GWAS to breast cancer

To further assess robustness of CC-GWAS, we applied CC-GWAS to BC case-control GWAS results of 61,282 cases + 45,494 controls (OncoArray sample in ref.[26]) vs. BC case-control GWAS results 46,785 cases + 42,892 controls (iCOGs sample in ref.[26]). Input parameters of CC-GWAS are the population prevalences[62], liability-scale heritabilities (h_l^2)[63–65], genetic correlation (r_g)[2], the intercept from cross-trait LD score regression[2] (used to model covariance of the error-terms), the sample size including overlap of controls ($N_{overlap} = 0$; also used to model covariance of the error-terms; see below), and expectation of the number of independent causal SNPs (m). The number of independent causal SNPs was set at $m = 7,500$[66] (see below for a detailed discussion of the assumed number of causal SNPs in applications of CC-GWAS). The resulting CC-GWAS$_{OLS}$ weights and CC-GWAS$_{Exact}$ weights are reported in Supplementary Table 9.

Application of CC-GWAS to psychiatric and other empirical data sets

Input parameters of CC-GWAS are the population prevalences (K), liability-scale heritabilities (h_l^2), genetic correlation (r_g), the intercept from cross-trait LD score regression[2] (used to model covariance of the error-terms), the sample sizes including overlap of controls (also used to model covariance of the error-terms; see below), and expectation of the number of independent causal SNPs (m; see below). Prevalences are displayed in Table 1 and Supplementary Table 25 and were based on ref.[67] for the eight psychiatric disorders, on ref. [68] for UC and CD, and ref.[53] for RA. Heritabilities were assessed with stratified LD score regression based on the baseline LD v2.0 model[63–65], and transposed to liability-scale[25,57]. Genetic correlations were estimated with cross-trait LD score regression[2]. The number of causal SNPs was set at $m = 10,000$ for the psychiatric disorders, and $m = 1,000$ for CD, UC and RA based on ref.[27]. CC-GWAS estimates the covariance of case-control error terms based on (i) the intercept of cross-trait LD score regression[7]) and (ii) sample overlap of controls (see Supplementary Note and Supplementary Table 39). CC-GWAS conservatively uses the minimum of these to estimates, because overestimation of the covariance of error terms will underestimate the standard error of CC-GWAS results thereby risking increased false positive rate. Based on the listed input parameters, CC-GWAS (software provided in R[69]) was applied one disorder pair at a time.

CC-GWAS results were clumped in line with ref.[16] using 1000 Genomes data[28] as LD reference panel with Plink 1.9[70] (--clump-p1 5e-8 --clump-p2 5e-8 --clump-r2 0.1 --clump-kb 3000) (Supplementary Table 12). Loci within 250kb of each other after the first clumping step were collapsed. We defined *CC-GWAS-specific* loci as loci for which none of the genome-wide significant SNPs have an $r^2 > 0.8$ with any of the genome-wide significant SNPs in the input case-control GWAS results (Supplementary Table 12). We chose this value because we think it is unlikely that a CC-GWAS locus would statistically result from a significant case-control locus for which all significant SNPs have r^2 0.8 with all significant SNPs in the CC-GWAS locus. An overview of the number of CC-GWAS loci is given in Table 1 and Supplementary Table 14, and details are reported in Table 2 and Supplementary Table 13. Secondary analyses are described in the Supplementary Note and Supplementary Table 18.

The assumed number of causal SNPs in applications of CC-GWAS

Our primary recommendation is to specify m based on published estimates of genome-wide polygenicity, such as the effective number of independently associated causal SNPS[27] or the total number of independently associated SNPs[60,66,71,72]. These values generally range from 1,000 for relatively sparse traits (e.g. autoimmune diseases) to 10,000 for highly polygenic traits (e.g. psychiatric disorders). When estimates of genome-wide polygenicity are not available, our recommendation is to specify $m=1,000$ for traits that are expected to have relatively sparse architectures (e.g. autoimmune diseases), $m=10,000$ for traits that are expected to have highly polygenic architectures (e.g. psychiatric disorders), and $m=5,000$ for traits with no clear expectation. When comparing disorders with different levels of polygenicity, our recommendation is to specify m based on the expected average across both disorders. We note that misspecification of m may impact power, but does not affect type I error rate (Supplementary Note and Supplementary Figure 12).

SMR and HEIDI analyses

We used the SMR test for colocalization[30] to identify CC-GWAS loci with significant associations between gene expression effect sizes in *cis* and CC-GWAS$_{OLS}$ case-case effect sizes. We tested cis-eQTL effects in 13 GTEx v7 brain tissues[31] (Amygdala, Anterior cingulate cortex, Caudate basal ganglia, Cerebellar Hemisphere, Cerebellum, Cortex, Frontal Cortex, Hippocampus, Hypothalamus, Nucleus accumbens basal ganglia, Putamen basal ganglia, Spinal cord cervical c-1, and Substantia nigra), and a meta-analysis of eQTL effects in brain tissues[32]. In line with standard application of SMR[30], we tested probes of genes with significant eQTL associations, with the lead eQTL SNP within 1MB of the lead CC-GWAS SNP. SMR analyses were performed on 2MB cis windows around the tested probe. The threshold of significance was adjusted per tested disorder-pair by dividing 0.05 by the respective number of probes tested (Supplementary Table 23). We used the HEIDI test for heterogeneity[30] to exclude loci with evidence of linkage effects ($P < 0.05$).

Replication analyses

Of the eight psychiatric disorders, only SCZ and MDD had sufficient sample size to perform replication analyses of the SCZ vs. MDD results based on publicly available GWAS results of subsets of the data[15,51]. We used CC-GWAS discovery results of additional analyses of SCZ vs. MDD based on GWAS results from Ripke et al.[15] and Wray et al.[51] (Supplementary Table 26). To obtain independent replication data, we applied MetaSubtract[54] separately for SCZ (results (i) Pardinas et al.[16] – results (ii) Ripke et al.[15]) and for MDD (results (i) Howard et al.[18] – results (ii) Wray et al.[51]; see Supplementary Note). For further replication analyses, we used CC-GWAS discovery results from the three comparisons of CD[52], UC[52] and RA[53]. For CD[52] and UC[52], we also applied MetaSubtract[54] to obtain independent discovery and replication results (Supplementary Note and Supplementary Table 26). In a final set of secondary replication analyses, we sought to perform replication analyses using independent replication data that was obtained without requiring the use of MetaSubtract, and we focused on 6 comparisons of each pair of 4 disorders[26,51,53,61,73] (of low biological interest, but useful for assessing the robustness of the CC-GWAS method; see Supplementary Note).

For replication analyses, we computed CC-GWAS$_{OLS}$ $A1B1$ effects based on the CC-GWAS$_{OLS}$ weights from the respective discovery results (Supplementary Table 26). (We applied CC-GWAS$_{OLS}$ weights from the discovery analyses rather than re-estimating the CC-GWAS$_{OLS}$ weights based on the replication GWAS results, because CC-GWAS$_{OLS}$ weights are sample size dependent.) We used the replication slope (based on a regression of replication vs. discovery effect sizes[55]) to assess the level of replication for all CC-GWAS loci and for the CC-GWAS-specific loci separately. We note that CC-GWAS-specific loci are expected to have smaller case-case effect sizes (because the respective case-control effect sizes are not significant per definition) than the remaining CC-GWAS loci. Thus, CC-GWAS-specific loci are expected to be more susceptible to winner's curse[56] and attain a lower replication slope than the remaining CC-GWAS loci (see Supplementary Note).

Data availability

CC-GWAS results generated in the present study for 8 psychiatric disorders and 3 autoimmune diseases are available for public download at https://data.broadinstitute.org/alkesgroup/CC-GWAS/. GWAS results for breast cancer are available at http://bcac.ccge.medschl.cam.ac.uk/bcacdata/. GWAS results for attention deficit/hyperactivity disorder, anorexia nervosa, autism spectrum disorder, bipolar disorder, major depressive disorder (Wray 2018), obsessive–compulsive disorder, schizophrenia (Ripke 2014), schizophrenia vs bipolar disorder, and Tourette's syndrome and other tic disorders are available at https://www.med.unc.edu/pgc/results-and-downloads/. GWAS results for major depressive disorder (Howard 2019) are available at https://datashare.is.ed.ac.uk/handle/10283/3203. GWAS results for schizophrenia (Pardinas 2018) are available at https://walters.psycm.cf.ac.uk/. GWAS results for Crohn's disorder and ulcerative colitis are available at https://www.ibdgenetics.org/downloads.html. GWAS results for rheumatoid arthritis are available at http://www.sg.med.osaka-u.ac.jp/tools.html. GWAS results for coronary artery disease are available at http://www.cardiogramplusc4d.org/data-downloads/. eQTL data of 13 GTEx v7 brain tissues and meta-analysis of eQTL effects in brain tissues are available at https://cnsgenomics.com/software/smr/#DataResource. Access to the UK Biobank resource is available via application (http://www.ukbiobank.ac.uk).

Code availability

CC-GWAS software (R package) is available at https://github.com/wouterpeyrot/CCGWAS. R software is available at https://www.r-project.org/ (version 3.5.1 was used). LDSC software is available at https://github.com/bulik/ldsc (version 1.0.0 was used). SMR software is available at https://cnsgenomics.com/software/smr/ (version 1.02 was used). PLINK1.9 software is available at www.cog-genomics.org/plink/1.9/ (version v1.90b6.7 was used).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References (main text only)

1. Lee SH et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat. Genet. 45, 984–94 (2013). [PubMed: 23933821]

2. Bulik-Sullivan B et al. An atlas of genetic correlations across human diseases and traits. Nat. Genet. (2015). doi:10.1038/ng.3406

3. Lee PH et al. Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. Cell 179, 1469–1482.e11 (2019). [PubMed: 31835028]

4. Ruderfer DM et al. Genomic Dissection of Bipolar Disorder and Schizophrenia, Including 28 Subphenotypes. Cell 173, 1705–1715.e16 (2018). [PubMed: 29906448]

5. Pasaniuc B & Price AL Dissecting the genetics of complex traits using summary association statistics. Nat. Rev. Genet. 18, 117–127 (2017). [PubMed: 27840428]

6. Lin DY & Sullivan PF Meta-Analysis of Genome-wide Association Studies with Overlapping Subjects. Am. J. Hum. Genet. 85, (2009).

7. Turley P et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. Nat. Genet. 50, 229–237 (2018). [PubMed: 29292387]

8. Qi G & Chatterjee N Heritability informed power optimization (HIPO) leads to enhanced detection of genetic associations across multiple traits. PLoS Genet. 14, e1007549 (2018). [PubMed: 30289880]

9. Baselmans BML et al. Multivariate genome-wide analyses of the well-being spectrum. Nat. Genet. 51, 445–451 (2019). [PubMed: 30643256]

10. Nieuwboer HA, Pool R, Dolan CV, Boomsma DI & Nivard MG GWIS: Genome-Wide Inferred Statistics for Functions of Multiple Phenotypes. Am. J. Hum. Genet. 99, 917–927 (2016). [PubMed: 27616482]

11. Bhattacharjee S et al. A subset-based approach improves power and interpretation for the combined analysis of genetic association studies of heterogeneous traits. Am. J. Hum. Genet. 90, 821–35 (2012). [PubMed: 22560090]

12. Han B & Eskin E Interpreting meta-analyses of genome-wide association studies. PLoS Genet. 8, e1002555 (2012). [PubMed: 22396665]

13. Zhu Z et al. Causal associations between risk factors and common diseases inferred from GWAS summary data. Nat. Commun. 9, 224 (2018). [PubMed: 29335400]

14. Byrne EM et al. Conditional GWAS analysis identifies putative disorder-specific SNPs for psychiatric disorders. Mol. Psychiatry (2020). doi:10.1038/s41380-020-0705-9

15. Ripke S et al. Biological insights from 108 schizophrenia-associated genetic loci. Nature 511, 421–7 (2014). [PubMed: 25056061]

16. Pardiñas AF et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. Nat. Genet. 50, 381–389 (2018). [PubMed: 29483656]

17. Stahl EA et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. Nat. Genet. 51, 793–803 (2019). [PubMed: 31043756]

18. Howard DM et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. Nat. Neurosci. 22, 343–352 (2019). [PubMed: 30718901]

19. Demontis D et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nat. Genet. 51, 63–75 (2019). [PubMed: 30478444]

20. Watson HJ et al. Genome-wide association study identifies eight risk loci and implicates metabo-psychiatric origins for anorexia nervosa. Nat. Genet. 51, 1207–1214 (2019). [PubMed: 31308545]

21. Grove J et al. Identification of common genetic risk variants for autism spectrum disorder. Nat. Genet. 51, 431–444 (2019). [PubMed: 30804558]

22. International Obsessive Compulsive Disorder Foundation Genetics Collaborative (IOCDF-GC) and OCD Collaborative Genetics Association Studies (OCGAS). Revealing the complex genetic architecture of obsessive-compulsive disorder using meta-analysis. Mol. Psychiatry 23, 1181–1188 (2018). [PubMed: 28761083]

23. Yu D et al. Interrogating the Genetic Determinants of Tourette's Syndrome and Other Tic Disorders Through Genome-Wide Association Studies. Am. J. Psychiatry 176, 217–227 (2019). [PubMed: 30818990]

24. Yang J, Wray NR & Visscher PM Comparing apples and oranges: equating the power of case-control and quantitative trait association studies. Genet. Epidemiol. 34, 254–7 (2010). [PubMed: 19918758]

25. Lee SH, Wray NR, Goddard ME & Visscher PM Estimating Missing Heritability for Disease from Genome-wide Association Studies. Am. J. Hum. Genet. 88, 294–305 (2011). [PubMed: 21376301]

26. Michailidou K et al. Association analysis identifies 65 new breast cancer risk loci. Nature 551, 92–94 (2017). [PubMed: 29059683]

27. O'Connor LJ et al. Extreme Polygenicity of Complex Traits Is Explained by Negative Selection. Am. J. Hum. Genet. 105, 456–476 (2019). [PubMed: 31402091]

28. 1000 Genomes Project Consortium et al. A global reference for human genetic variation. Nature 526, 68–74 (2015). [PubMed: 26432245]

29. Buniello A et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. 47, D1005–D1012 (2019). [PubMed: 30445434]

30. Zhu Z et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat. Genet. 48, (2016).

31. GTEx Consortium et al. Genetic effects on gene expression across human tissues. Nature 550, 204–213 (2017). [PubMed: 29022597]

32. Qi T et al. Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. Nat. Commun. 9, 2282 (2018). [PubMed: 29891976]

33. O'Leary NA et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 44, D733–45 (2016). [PubMed: 26553804]

34. Hoefs SJG et al. NDUFA2 complex I mutation leads to Leigh disease. Am. J. Hum. Genet. 82, 1306–15 (2008). [PubMed: 18513682]

35. Li Z et al. Genome-wide association analysis identifies 30 new susceptibility loci for schizophrenia. Nat. Genet. 49, 1576–1583 (2017). [PubMed: 28991256]

36. Ikeda M et al. Genome-Wide Association Study Detected Novel Susceptibility Genes for Schizophrenia and Shared Trans-Populations/Diseases Genetic Effect. Schizophr. Bull. 45, 824–834 (2019). [PubMed: 30285260]

37. Lam M et al. Comparative genetic architectures of schizophrenia in East Asian and European populations. Nat. Genet. 51, 1670–1678 (2019). [PubMed: 31740837]

38. Lam M et al. Pleiotropic Meta-Analysis of Cognition, Education, and Schizophrenia Differentiates Roles of Early Neurodevelopmental and Adult Synaptic Pathways. Am. J. Hum. Genet. 105, 334–350 (2019). [PubMed: 31374203]

39. Nagel M et al. Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. Nat. Genet. 50, 920–927 (2018). [PubMed: 29942085]

40. Lee JJ et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat. Genet. 50, 1112–1121 (2018). [PubMed: 30038396]

41. Savage JE et al. Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. Nat. Genet. 50, 912–919 (2018). [PubMed: 29942086]

42. Giri A et al. Trans-ethnic association study of blood pressure determinants in over 750,000 individuals. Nat. Genet. 51, 51–62 (2019). [PubMed: 30578418]

43. Evangelou E et al. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. Nat. Genet. 50, 1412–1425 (2018). [PubMed: 30224653]

44. The UniProt Consortium. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 45, D158–D169 (2017). [PubMed: 27899622]

45. Okbay A et al. Genome-wide association study identifies 74 loci associated with educational attainment. Nature 533, 539–542 (2016). [PubMed: 27225129]

46. Hill WD et al. A combined analysis of genetically correlated traits identifies 187 loci and a role for neurogenesis and myelination in intelligence. Mol. Psychiatry 24, 169–181 (2019). [PubMed: 29326435]

47. Davies G et al. Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. Nat. Commun. 9, 2098 (2018). [PubMed: 29844566]

48. Moore DL, Apara A & Goldberg JL Krüppel-like transcription factors in the nervous system: novel players in neurite outgrowth and axon regeneration. Mol. Cell. Neurosci. 47, 233–43 (2011). [PubMed: 21635952]

49. Sekar A et al. Schizophrenia risk from complex variation of complement component 4. Nature 530, 177–83 (2016). [PubMed: 26814963]

50. Yanagi M et al. Expression of Kruppel-like factor 5 gene in human brain and association of the gene with the susceptibility to schizophrenia. Schizophr. Res. 100, 291–301 (2008). [PubMed: 18226501]

51. Wray NR et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nat. Genet. 50, 668–681 (2018). [PubMed: 29700475]

52. Liu JZ et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nat. Genet. 47, 979–986 (2015). [PubMed: 26192919]

53. Okada Y et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature 506, 376–81 (2014). [PubMed: 24390342]

54. Nolte IM et al. Missing heritability: is the gap closing? An analysis of 32 complex traits in the Lifelines Cohort Study. Eur. J. Hum. Genet. 25, 877–885 (2017). [PubMed: 28401901]

55. Marigorta UM & Navarro A High trans-ethnic replicability of GWAS results implies common causal variants. PLoS Genet. 9, e1003566 (2013). [PubMed: 23785302]

56. Palmer C & Pe'er I Statistical correction of the Winner's Curse explains replication variability in quantitative trait genome-wide association studies. PLOS Genet. 13, e1006916 (2017). [PubMed: 28715421]

References (Method-only)

57. Golan D, Lander ES & Rosset S Measuring missing heritability: Inferring the contribution of common variants. Proc. Natl. Acad. Sci. U. S. A. 111, E5272–81 (2014). [PubMed: 25422463]

58. Bhatia G, Patterson N, Sankararaman S & Price AL Estimating and interpreting FST: The impact of rare variants. Genome Res. 23, 1514–1521 (2013). [PubMed: 23861382]

59. Lloyd-Jones LR, Robinson MR, Yang J & Visscher PM Transformation of Summary Statistics from Linear Mixed Model Association on All-or-None Traits to Odds Ratio. Genetics 208, 1397–1408 (2018). [PubMed: 29429966]

60. Frei O et al. Bivariate causal mixture model quantifies polygenic overlap between complex traits beyond genetic correlation. Nat. Commun. 10, 2417 (2019). [PubMed: 31160569]

61. Bycroft C et al. The UK Biobank resource with deep phenotyping and genomic data. Nature 562, (2018).

62. Siegel R, Ma J, Zou Z & Jemal A Cancer statistics, 2014. CA. Cancer J. Clin. 64, 9–29 (2014). [PubMed: 24399786]

63. Finucane HK et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet. (2015). doi:10.1038/ng.3404

64. Gazal S et al. Linkage disequilibrium–dependent architecture of human complex traits shows action of negative selection. Nat. Genet. 49, 1421–1427 (2017). [PubMed: 28892061]

65. Gazal S, Marquez-Luna C, Finucane HK & Price AL Reconciling S-LDSC and LDAK functional enrichment estimates. Nat. Genet. 51, 1202–1204 (2019). [PubMed: 31285579]

66. Zhang YD et al. Assessment of polygenic architecture and risk prediction based on common variants across fourteen cancers. Nat. Commun. (2020). doi:10.1038/s41467-020-16483-3

67. Sullivan PF & Geschwind DH Defining the Genetic, Genomic, Cellular, and Diagnostic Architectures of Psychiatric Disorders. Cell 177, 162–183 (2019). [PubMed: 30901538]

68. Molodecky NA et al. Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. Gastroenterology 142, 46–54.e42; quiz e30 (2012). [PubMed: 22001864]

69. R Core Team. R: A Language and Environment for Statistical Computing. (2018).

70. Chang CC et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4, 7 (2015). [PubMed: 25722852]

71. Zhang Y, Qi G, Park J-H & Chatterjee N Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits. Nat. Genet. 50, 1318–1326 (2018). [PubMed: 30104760]

72. Zeng J et al. Signatures of negative selection in the genetic architecture of human complex traits. Nat. Genet. 50, 746–753 (2018). [PubMed: 29662166]

73. Schunkert H et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. Nat. Genet. 43, (2011).
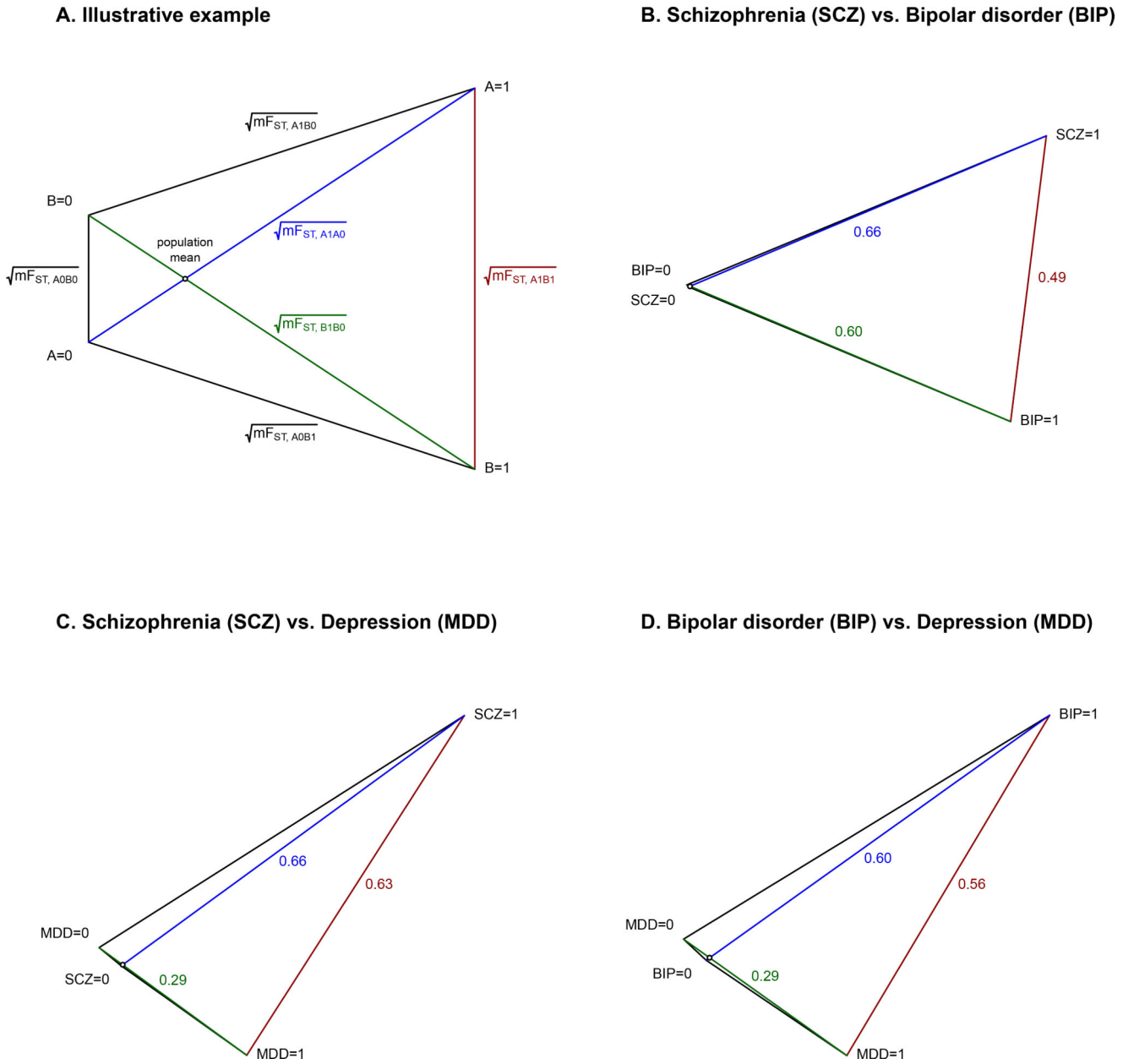
A. Illustrative example

B. Schizophrenia (SCZ) vs. Bipolar disorder (BIP)

C. Schizophrenia (SCZ) vs. Depression (MDD)

D. Bipolar disorder (BIP) vs. Depression (MDD)

Figure 1. Genetic distance between cases and/or controls of SCZ, BIP and MDD.
We report genetic distances for (A) an illustrative example, (B) SCZ vs. BIP, (C) SCZ vs. MDD and (D) SCZ vs. BIP. Genetic distances are displayed as $\sqrt{m * F_{ST, causal}}$, derived based on the respective population prevalences, SNP-based heritabilities and genetic correlations reported in Table 1 (m, number of independent causal variants; see Methods). Approximate standard errors of $m * F_{ST,causal,A1B1}$ are 0.04 for SCZ vs. BIP, 0.02 for SCZ vs. MDD and 0.03 for BIP vs. MDD (see Methods). For SCZ and BIP, despite the large genetic correlation ($r_g = 0.7$), the genetic distance between SCZ cases and BIP cases is only slightly smaller ($\sqrt{m * F_{ST, causal}} = 0.49$) than the case-control distances for SCZ (0.66) and BIP (0.60), because of the doubly strong ascertainment (due to low disorder prevalences) in SCZ cases and BIP cases and because a genetic correlation of 0.7 is still considerably

smaller than a genetic correlation of e.g. 0.9 (Supplementary Figure 20). For SCZ and MDD ($r_g = 0.31$), the genetic distance between MDD cases and SCZ cases (0.63) is larger than for MDD case-control (0.29) (Panel C) owing to the larger prevalence and lower heritability of MDD. For MDD and BIP ($r_g = 0.33$), genetic distances are similar to MDD and SCZ (Panel D). Numerical results are reported in Supplementary Table 11.
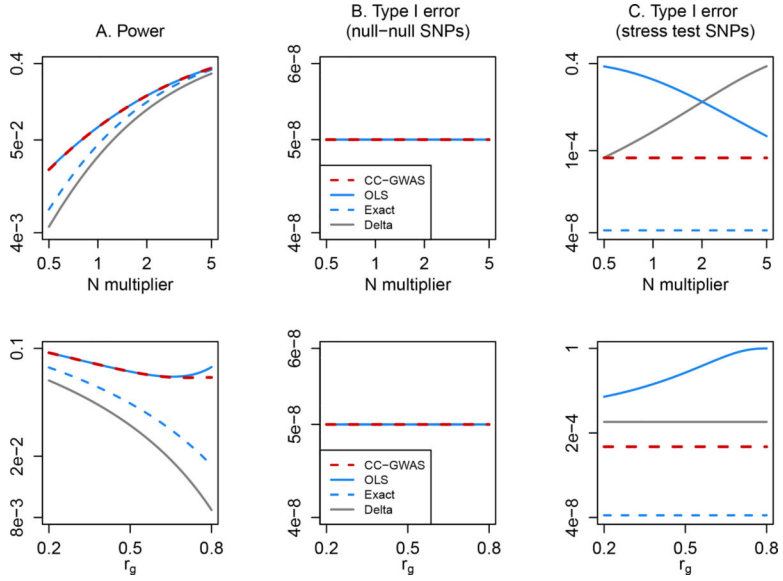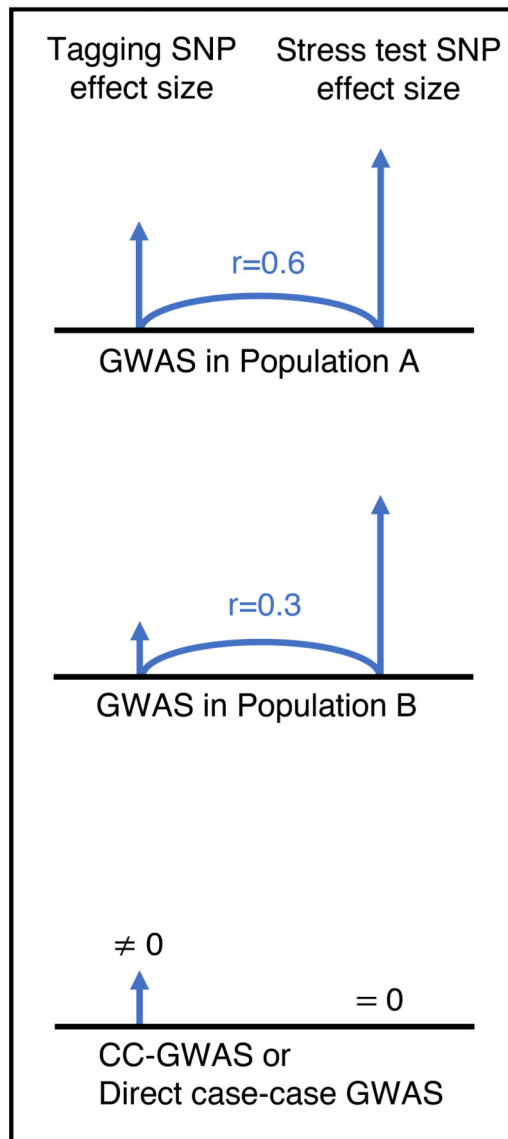
Figure 2. Power and type I error of CC-GWAS.
We report (A) the power to detect SNPs with effect sizes following a bivariate normal distribution, (B) the type I error rate for loci with no effect on A1A0 or B1B0 ("null-null" SNPs) and (C) the type I error rate for SNPs with the same allele frequency in A1 vs. B1 that explain 0.10% of variance in A1 vs. A0 and 0.29% of variance in B1 vs. B0 ("stress test" SNPs), for each of four methods: CC-GWAS, the CC-GWAS$_{OLS}$ component, the CC-GWAS$_{Exact}$ component, and a naïve Delta method (see text). Default parameter settings are: h^2=0.2, prevalence K=0.01, and sample size 100,000 cases + 100,000 controls for disorder A; liability-scale h^2=0.1, prevalence K=0.15, and sample size 100,000 cases + 100,000 controls for disorder B; m=5,000 causal SNPs for each disorder; and genetic correlation r_g=0.5 between disorders. Numerical results of these analytical computations are reported in Supplementary Table 1, and confirmed with simulation results in Supplementary Table 2.

A. Illustrative example

B. Simulation results
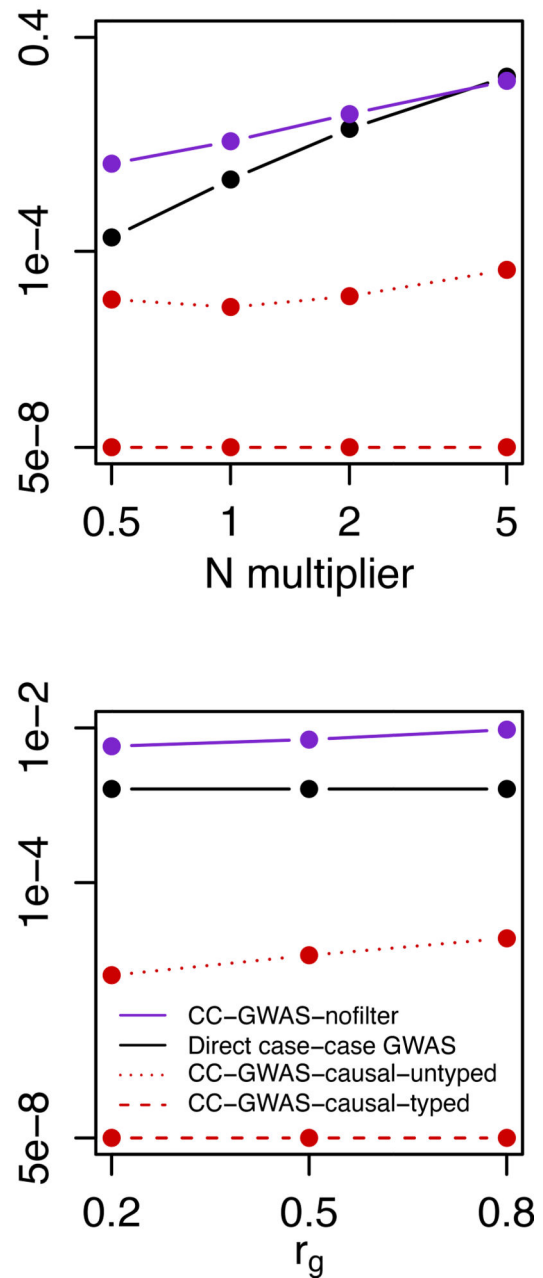
Figure 3. Type I error of CC-GWAS due to differential tagging of a causal stress test SNP.
(A) Illustrative example of how differential tagging of a causal stress test SNP can lead to type I error. (B) Simulation results of type I error due to differential tagging. We report the per-locus type I error rate, defined as the number of loci with at least one genome-wide significant tagging SNP divided by the number of loci tested, for each of four methods/scenarios: CC-GWAS, causal stress test SNP genotyped/imputed (denoted CC-GWAS-causal-typed); CC-GWAS, causal stress test SNP not genotyped/imputed (denoted CC-GWAS-causal-untyped); CC-GWAS, no filter; and Direct case-case GWAS (see text).

Default parameter settings are: h^2=0.2, prevalence K=0.01, and sample size 100,000 cases + 100,000 controls for disorder A; liability-scale h^2=0.1, prevalence K=0.15, and sample size 100,000 cases + 100,000 controls for disorder B; m=5,000 causal SNPs for each disorder; and genetic correlation r_g=0.5 between disorders. Per-locus type I error rates <5×10^{-8} were truncated to 5×10^{-8} for visualization purposes. All simulation standard errors were 0 for CC-GWAS-causal-typed (zero false positives across all simulations performed); 2.7×10^{-5} for CC-GWAS-causal-untyped; 1.7×10^{-3} for CC-GWAS, no filter; and 2.1×10^{-3} for Direct case-case GWAS. Numerical results are reported in Supplementary Table 7.
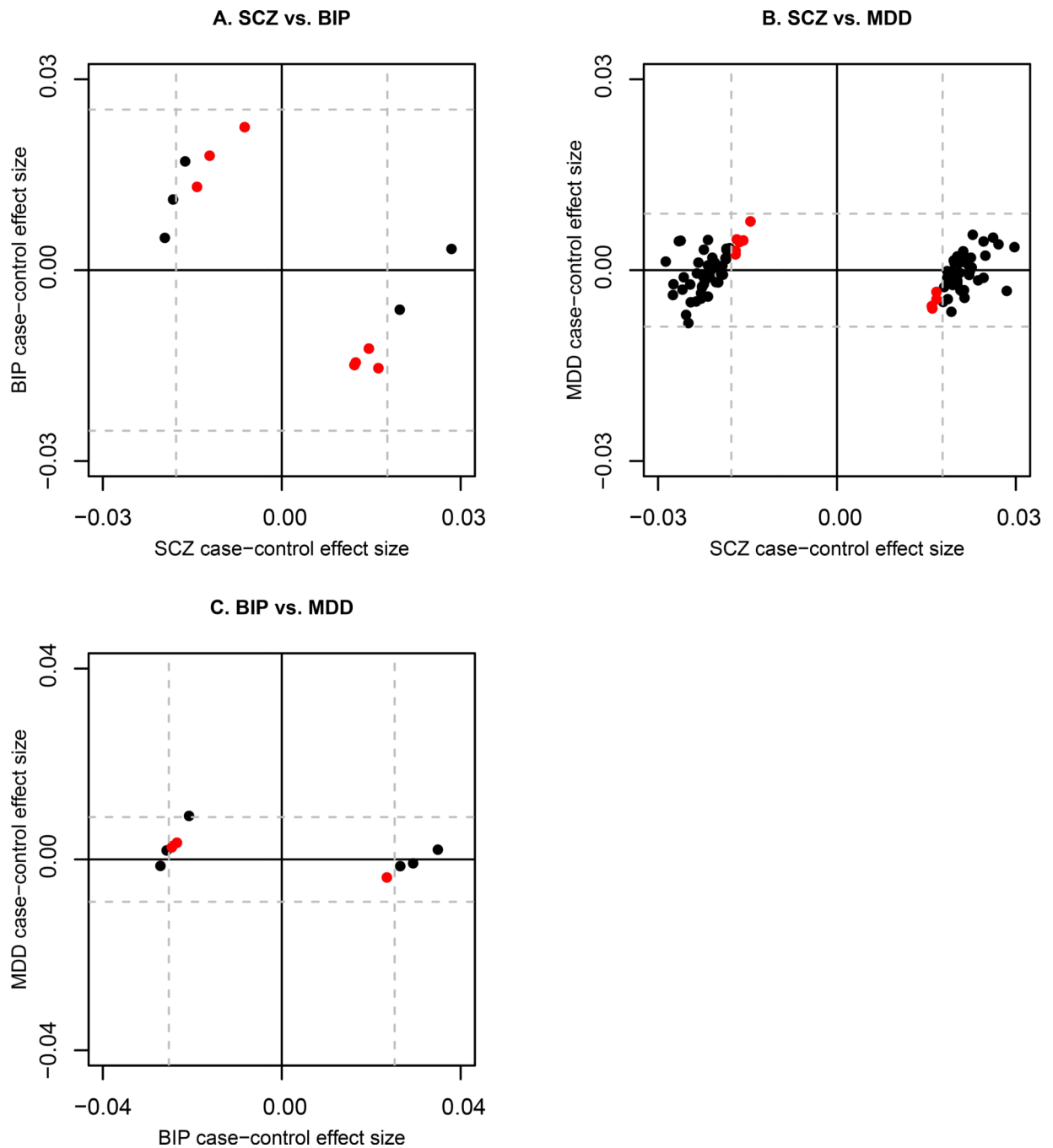
Figure 4. Case-control effect sizes at CC-GWAS loci for SCZ, BIP and MDD.
We report the respective case-control effect sizes for lead SNPs at CC-GWAS loci for (A) SCZ vs. BIP, (B) SCZ vs. MDD and (C) BIP vs. MDD. Effect sizes are reported on the standardized observed scale based on 50/50 case-control ascertainment. Red points denote CC-GWAS-specific loci, and black points denote remaining loci. Dashed lines denote effect-size thresholds for genome-wide significance. All red points (denoting lead SNPs for CC-GWAS-specific loci) lie inside the dashed lines for both disorders; in panel A, one black point (denoting the lead SNP for a CC-GWAS locus that is not CC-GWAS-specific) lies

inside the dashed lines for both SCZ and BIP, because the lead SNP is not genome-wide significant for SCZ but is in LD with a SNP that is genome-wide significant for SCZ. Numerical results are reported in Supplementary Table 13. SCZ, schizophrenia; BIP, bipolar disorder; MDD, major depressive disorder.
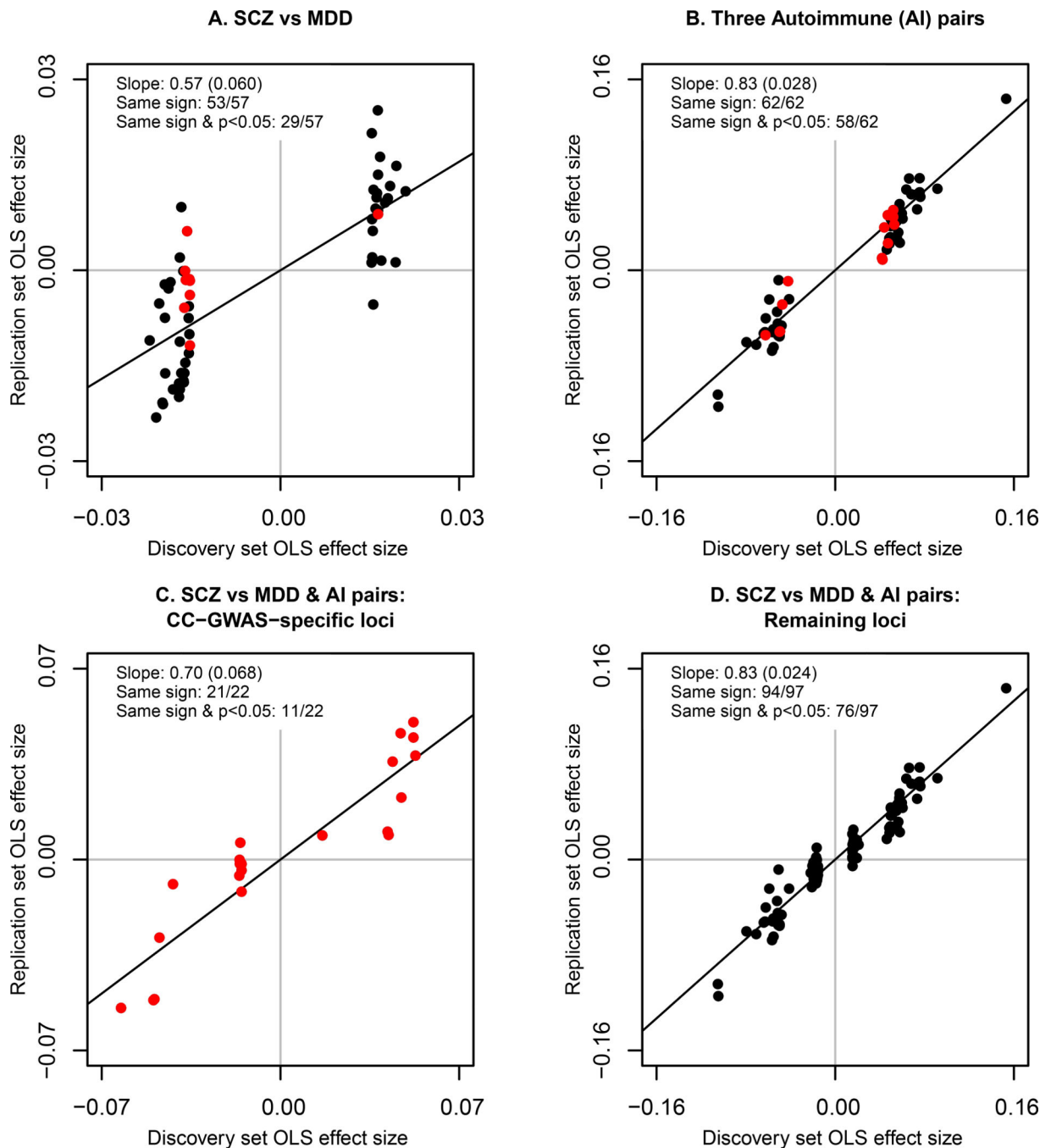
Figure 5. Independent replication of CC-GWAS results.
We report replication case-case CC-GWAS$_{OLS}$ effect sizes vs. discovery case-case CC-GWAS$_{OLS}$ effect sizes for (A) schizophrenia (SCZ) vs. major depressive disorder (MDD), (B) three autoimmune disorders, (C) SCZ vs. MDD and three autoimmune disorders, restricting to CC-GWAS-specific loci, and (D) SCZ vs. MDD and three autoimmune disorders, restricting to remaining loci. We also report regression slopes (SE in parentheses), effect sign concordance, and effect sign concordance together with replication $P_{OLS}<0.05$. Red points denote CC-GWAS-specific loci, and black points denote remaining loci.

Numerical results are reported in Supplementary Table 27, and corresponding case-control replication results are reported in Supplementary Figure 15 and Supplementary Table 28.

Table 1.

Summary of CC-GWAS results for schizophrenia, bipolar disorder and major depressive disorder.

For each pair of schizophrenia (SCZ)[16], bipolar disorder (BIP)[17] and major depressive disorder (MDD)[18], we report the case-control sample sizes, #SNPs, the most likely prevalence (K)[67], liability-scale heritability estimated using stratified LD score regression[63–65] (h^2), genetic correlation estimated using cross-trait LD score regression[2] (r_g), CC-GWAS$_{OLS}$ weights (based on the most likely prevalences), number of independent genome-wide significant loci for each case-control comparison, number of independent genome-significant CC-GWAS loci, and number of independent genome-significant CC-GWAS loci that are CC-GWAS-specific. CC-GWAS$_{Exact}$ weights are equal to $(1 - K_A)$ for disorder A and $-(1 - K_B)$ for disorder B. CC-GWAS reports a SNP as statistically significant if it achieves $P < 5 * 10^{-8}$ using CC-GWAS$_{OLS}$ weights *and* $P < 10^{-4}$ using CC-GWAS$_{Exact}$ weights. We specified a range of prevalences to the CC-GWAS$_{Exact}$ component for SCZ $(0.4\%-1.0\%)$[16,67], BIP $(0.5\%-2.0\%)$[17], and MDD $(16\%-30\%)$[18,67] (yielding $2 \times 2 = 4$ CC-GWAS$_{Exact}$ p-values per comparison, all required to be $< 10^{-4}$). Statistical tests are two-sided.

| | | | | A1A0 | | B1B0 | | | Number of significant independent loci | | | |
| | | | | | | | | | | | | CC-GWAS |
A1A0 (N case/ N control)	B1B0 (N case/ N control)	# SNPs	K (%)	h^2	K (%)	h^2	r_g	OLS weights	A1A0	B1B0	all	CC-GWAS specific
SCZ (40,675/64,643)	BIP (20,352/31,358)	4,548,414	0.40	0.20	1.00	0.20	0.70	0.55/−0.43	139	15	**12**	7
SCZ (40,675/64,643)	MDD (170,756/329,443)	4,483,387	0.40	0.20	16.00	0.10	0.31	0.77/−0.51	139	50	**99**	10
BIP (20,352/31,358)	MDD (170,756/329,443)	6,265,453	1.00	0.20	16.00	0.10	0.33	0.58/−0.43	14	53	**10**	4

Table 2.
List of 21 CC-GWAS-specific loci for SCZ, BIP and MDD.

For each CC-GWAS-specific locus, we report the lead CC-GWAS SNP and its chromosome, physical position, and reference allele frequency, the locus name, the respective case-control effect sizes and p-values, and the CC-GWAS$_{OLS}$ case-case effect size and p-value. Effect sizes are reported on the standardized observed scale based on 50/50 case-control ascertainment.

Disorder							A1A0		B1B0		A1B1 (OLS)	
A	B	SNP	Chr	Position	Freq	Locus name	Beta	P	Beta	P	Beta	P
SCZ	BIP	rs9866687[a]	3	94,828,190	0.44	LINC00879	1.24e-02	1.38e-04	−1.45e-02	1.70e-03	1.30e-02	4.05e-08
SCZ	BIP	rs7790864[a]	7	28,478,625	0.38	CREB5	1.46e-02	7.18e-06	−1.23e-02	7.93e-03	1.32e-02	2.18e-08
SCZ	BIP	rs12554512	9	23,352,293	0.43	ELAVL2	−6.22e-03	5.54e-02	2.25e-02	1.28e-06	−1.30e-02	4.06e-08
SCZ	BIP	rs3764002	12	108,618,630	0.26	WSCD2[c]	1.62e-02	6.05e-07	−1.54e-02	9.04e-04	1.55e-02	6.33e-11
SCZ	BIP	rs9319540[a]	16	79,458,022	0.58	MAF	1.22e-02	1.84e-04	−1.49e-02	1.26e-03	1.30e-02	3.67e-08
SCZ	BIP	rs1054972	19	1,852,582	0.2	KLF16[c]	−1.42e-02	1.32e-05	1.31e-02	4.74e-03	−1.33e-02	1.75e-08
SCZ	BIP	rs11696888	20	47,753,265	0.43	CSE1L[b]	−1.21e-02	1.94e-04	1.80e-02	1.05e-04	−1.43e-02	1.39e-09
SCZ	MDD	rs2471403	2	48,490,508	0.48	FOXN2[b]	−1.70e-02	1.78e-07	3.08e-03	6.21e-02	−1.47e-02	2.34e-08
SCZ	MDD	rs16846133[a]	2	212,289,728	0.31	ERBB4	−1.63e-02	5.68e-07	4.43e-03	5.84e-03	−1.48e-02	1.71e-08
SCZ	MDD	rs2563297	5	140,097,072	0.44	PCDHA7[b]	1.60e-02	8.76e-07	−6.00e-03	3.73e-04	1.54e-02	5.25e-09
SCZ	MDD	rs113113059	6	43,160,375	0.19	CUL9[b]	−1.68e-02	2.37e-07	4.84e-03	2.93e-03	−1.55e-02	4.11e-09
SCZ	MDD	rs2944833	7	71,774,496	0.57	CALN1[b]	−1.70e-02	1.80e-07	2.52e-03	1.21e-01	−1.44e-02	4.22e-08
SCZ	MDD	rs71523422[a]	8	31,445,336	0.08	NRG1	−1.57e-02	1.41e-06	4.69e-03	3.82e-03	−1.45e-02	3.38e-08
SCZ	MDD	rs10967586[a]	9	26,895,808	0.13	CAAP1	1.67e-02	2.87e-07	−4.57e-03	4.60e-03	1.52e-02	6.94e-09
SCZ	MDD	rs17731	10	3,821,561	0.35	KLF6[c]	1.67e-02	2.86e-07	−3.39e-03	3.89e-02	1.46e-02	2.64e-08
SCZ	MDD	rs34232444	19	4,965,404	0.3	UHRF1	−1.45e-02	8.70e-06	7.66e-03	2.56e-06	−1.51e-02	9.92e-09
SCZ	MDD	rs8137258[a]	22	20,135,961	0.22	ZDHHC8	1.59e-02	9.87e-07	−5.65e-03	4.50e-04	1.52e-02	7.82e-09
BIP	MDD	rs28565152	5	7,542,911	0.25	ADCY2	2.35e-02	3.83e-07	−3.77e-03	2.23e-02	1.53e-02	2.79e-08
BIP	MDD	rs12538191[a]	7	44,980,824	0.24	SNHG15[b]	−2.44e-02	1.46e-07	2.81e-03	8.21e-02	−1.54e-02	2.36e-08
BIP	MDD	rs4447398	15	42,904,904	0.88	LRRC57[b]	−2.46e-02	1.10e-07	2.54e-03	1.22e-01	−1.54e-02	2.28e-08
BIP	MDD	rs11908600	20	43,633,418	0.3	STK4[b]	−2.34e-02	4.26e-07	3.53e-03	2.90e-02	−1.51e-02	4.16e-08

[a] denotes loci that have not been reported previously[29].

[b] denotes locus names based on (most) significant SMR results.

[c] denotes locus names based on exonic lead SNPs. Remaining locus names are based on nearest gene, and do not refer to any inferred biological function. Case-case effect sizes and p-values for the CC-GWAS$_{Exact}$ component are reported in Supplementary Table 13. SCZ, schizophrenia; BIP, bipolar disorder; MDD, major depressive disorder.

Table 3.
Summary of CC-GWAS results for eight psychiatric disorders.

For each pair of disorders, we report the genetic correlation estimated using cross-trait LD score regression[2] (r_g) (lower left) and the number of independent genome-significant CC-GWAS loci (number of CC-GWAS-specific loci in parentheses) (upper right). The respective case-control heritabilities, prevalences and GWAS sample sizes (input parameters of CC-GWAS) are presented in Supplementary Table 25 and its legend. The CC-GWAS$_{OLS}$ weights and number of SNPs tested are reported in Supplementary Table 14. CC-GWAS reports a SNP as statistically significant if it achieves $P < 5 * 10^{-8}$ using CC-GWAS$_{OLS}$ weights *and* $P < 10^{-4}$ using CC-GWAS$_{Exact}$ weights. Statistical tests are two-sided. SCZ, schizophrenia; BIP, bipolar disorder; MDD, major depressive disorder; ADHD, attention deficit/hyperactivity disorder; AN, anorexia nervosa; ASD, autism spectrum disorder; OCD, obsessive–compulsive disorder; TS, Tourette's Syndrome and Other Tic Disorders.

rg\# loci	SCZ	BIP	MDD	ADHD	ANO	ASD	OCD	TS
SCZ	-	12 (7)	99 (10)	43 (14)	41 (5)	40 (10)	0 (0)	13 (4)
BIP	0.70	-	10 (4)	8 (6)	5 (2)	3 (0)	1 (1)	5 (3)
MDD	0.31	0.33	-	9 (2)	6 (1)	3 (2)	0 (0)	0 (0)
ADHD	0.16	0.18	0.44	-	4 (3)	1 (0)	2 (2)	2 (2)
AN	0.26	0.10	0.28	0.01	-	1 (1)	0 (0)	2 (1)
ASD	0.25	0.17	0.34	0.37	0.11	-	1 (1)	1 (1)
OCD	0.32	0.27	0.25	-0.20	0.42	0.10	-	1 (1)
TS	0.11	0.08	0.23	0.19	0.08	0.16	0.50	-