Conditional GWAS analysis to identify disorder-specific SNPs for psychiatric disorders

Enda M Byrne[1], **Zhihong Zhu**[1], **Ting Qi**[1], **Nathan G Skene**[2], **Julien Bryois**[3], **Antonio F Pardinas**[4], **Eli Stahl**[5], **Jordan W Smoller**[6,7,8,9], **Marcella Rietschel**[10], **Bipolar Working Group of the Psychiatric Genomics Consortium**, **Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium**, **Michael J Owen**[4], **James T.R. Walters**[4], **Michael C O'Donovan**[4], **John G McGrath**[11,12,13], **Jens Hjerling-Leffler**[2], **Patrick F Sullivan**[3,14,15], **Michael E Goddard**[16,17], **Peter M Visscher**[1,11], **Jian Yang**[1,11], **Naomi R Wray**[1,11]

[1]Institute for Molecular Biosciences, University of Queensland, Brisbane, Australia

[2]Laboratory of Molecular Neurobiology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden

[3]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

[4]MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, UK

[5]Icahn School of Medicine at Mount Sinai, New York, NY, USA

[6]Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA, USA

[7]Psychiatric and Neurodevelopmental Genetics Unit, Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA

[8]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA

[9]Department of Psychiatry, Harvard Medical School, Boston, MA, USA

[10]Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, Heidelberg

[11]Queensland Brain Institute, University of Queensland, Brisbane, Australia

[12]Queensland Centre for Mental Health Research, The Park Centre for Mental Health, Wacol, Australia

[13]National Centre for Register-Based Research, Aarhus University, Aarhus, Denmark

[14]Department of Genetics, University of North Carolina, Chapel Hill, NC 27599-7264, USA

[15]Department of Psychiatry, University of North Carolina, Chapel Hill, NC 27599-7264, USA

[16]Faculty of Veterinary and Agricultural Science, University of Melbourne, Parkville, Victoria, Australia

[17]Biosciences Research Division, Department of Economic Development, Jobs, Transport and Resources, Bundoora, Victoria, Australia

Abstract

Substantial genetic liability is shared across psychiatric disorders but less is known about risk variants that are specific to a given disorder. We used multi-trait conditional and joint analysis (mtCOJO) to adjust GWAS summary statistics of one disorder for the effects of genetically correlated traits to identify putative disorder-specific SNP associations. We applied mtCOJO to summary statistics for five psychiatric disorders from the Psychiatric Genomics Consortium – schizophrenia (SCZ), bipolar disorder (BIP), major depression (MD), attention-deficit hyperactivity disorder (ADHD) and autism (AUT). Most genome-wide significant variants for these disorders had evidence of pleiotropy (i.e., impact on multiple psychiatric disorders) and hence have reduced mtCOJO conditional effect sizes. However, subsets of genome-wide significant variants had larger conditional effect sizes consistent with disorder-specific effects: 15 of 130 genome-wide significant variants for schizophrenia, 5 of 40 for major depression, 3 of 11 for ADHD and 1 of 2 for autism. We show that decreased expression of *VPS29* in the brain may increase risk to SCZ only and increased expression of *CSE1L* is associated with SCZ and MD, but not with BIP. Likewise, decreased expression of *PCDHA7* in the brain is linked to increased risk of MD but decreased risk of SCZ and BIP.

Introduction

Pervasive sharing of genetic risk factors between common psychiatric disorders (i.e. pleiotropy) has now been unequivocally demonstrated from genome-wide association studies (GWAS), as quantified by estimates of genetic correlation (r_g) [1,2]. The r_g estimates are highest between schizophrenia and bipolar disorder (0.67, standard error (s.e.) = 0.03) but are > 0.15 for any combination of the five common disorders of schizophrenia (SCZ), bipolar disorder (BIP), ADHD, Major Depression (MD) and autism spectrum disorders (AUT) [2,3]. Cross-diagnosis analyses can leverage power to identify genetic risk loci shared across classical diagnostic boundaries [4] and can increase power for risk prediction of disorders in independent samples [5,6]. The shared genetic basis for psychiatric disorders contributes to an evidence base supporting a trans-diagnostic approach in clinical practice [7]. Nonetheless, traditional diagnostic classes reflect real symptom differences at patient presentation even though it can be difficult to classify some individuals given a high-degree of concurrent and longitudinal comorbidity. Since r_g estimates are higher between data sets of the same disorder than between data sets of different disorders [4,8] it implies some real biological basis to the classical diagnostic classes. Hence, a key question of importance in psychiatry is identification of genetic factors that are disorder specific rather than those shared across classical diagnostic groupings. Identifying such variants could aid in understanding the biological pathways that underlie the constellation of symptoms seen in each disorder.

One method for identifying disorder-specific variants is to conduct a case-case GWAS with cases of one disorder compared to cases of another. The SCZ/BIP working group of the Psychiatric Genomics Consortium (PGC) conducted an association analysis comparing in logistic regression SCZ (N=23,585) vs BIP (N=15,270) cases to identify variants specific to each disorder. The cases were matched on ancestry and genotyping platform, hence the sample sizes were smaller than those available for the disorder specific GWAS which limits the statistical power. Conducting such analysis requires access to the raw genotypes which is

not always feasible for all cohorts due to privacy laws. Methods that use summary statistics can utilise larger sample sizes without the need to provide access to raw daa to researchers. In addition, case-case GWAS can identify differences between pairs of disorders[9], but does not generalize to the multivariate space to identify SNPs primarily associated one disorder.

We conditioned the effect of SNPs estimated for one disorder on those of other disorders using multi-trait, conditional and joint analysis (mtCOJO)[10], a summary-statistics based method that accounts for overlap in samples contributing to the disorder specific GWAS. We report results from conditional analyses of 5 psychiatric disorders: SCZ, BIP, MD, ADHD and AUT using association summary statistics from meta-analyses conducted by the Psychiatric Genomics Consortium (PGC) including data from 23andMe. Each disorder is conditioned on the other four disorders in one model.

Methods

We applied the mtCOJO method as described in Zhu et al. [10]. This method approximates a conditional analysis where the effect of a SNP on a disease is conditioned upon the covariates of the disease, but only requires summary statistics as input. As an example, if we are interested in estimating the effect of a SNP (z) on risk to schizophrenia (y) accounting for the effect of a covarying factor such as bipolar (x), we condition upon the effect of bipolar on schizophrenia $\hat{b}_{xy}$, as estimated using Generalised Summary-based Mendelian Randomisation (GSMR). This can be extended to condition upon multiple covarying diseases so that the effect of the SNP on risk on the disorder of interest is estimated conditional upon the covariates on the disorder (see Supplementary Material for detailed description of the method).

To identify independent genome-wide significant SNPs for use as genetic instruments in mtCOJO analysis, each dataset was clumped to select independent genome-wide significant (GWS) SNPs ($p < 5 \times 10^{-8}$) using 7,762 unrelated individuals from the Atherosclerosis Risk In Community (ARIC) dataset[11], imputed to 1000Genomes Phase III as an LD reference sample. GWS SNPs more than 1MB apart or with an r^2 value < 0.05 were considered to be independent. GSMR accounts for any remaining LD between instruments. GSMR analysis with filtering to remove SNPs with outlier pleiotropic effects (compared to other GWS SNPs) using the HEIDI test [12] was performed with each disorder included both as an exposure and an outcome in combination with the other disorders. Owing to having fewer than 10 independent GWS SNPs, independent SNPs significant at $p < 10^{-7}$ were used for GSMR analysis with autism as the exposure variable. In order to compare the estimated effects of one disorder on another from MR, we derived a conversion of the estimated effects from GSMR to the liability scale (see Supplementary Material, Supplementary Figure 1).

We performed mtCOJO analysis (implemented in GCTA[13] (http://cnsgenomics.com/software/gcta/#mtCOJO) of 5 genetically correlated psychiatric disorders using the results from large genome-wide association studies from the Psychiatric GWAS consortium (Table 1), running the analysis in turn with each disorder as the outcome with the other disorders as covariates. A total of 5,275,400 SNPs with matching alleles that were in common across the 5 disorders were used for further analysis. Indels were excluded from the analysis.

For each disorder, SNP effects conditional upon the other disorders were calculated. Results were uploaded to FUMA for annotation [14]. Ranking SNPs according to the difference between the marginal and conditional effect sizes for each disorder is not necessarily meaningful because, for example, a SNP that has a low estimated marginal effect, so no effect on the outcome trait, will have a large conditional effect if the SNP has a large effect on the covariate traits. For the purposes of identifying which SNPs show evidence of disorder-specificity, we focus on presenting results for SNPs that were GWS with the outcome disorder in the original GWAS. We further estimated whether the difference between the conditional and raw effect size of each SNP was significant (Supplementary Material).

MAGMA gene-set analysis

MAGMA gene-set analysis[15] as implemented in FUMA was used to investigate which sets of biologically related genes show the strongest evidence of association in the conditional analyses.

Genetic Correlation

LD-score regression [16] was used to estimate the genetic correlation between the conditional and unadjusted GWAS results.

Summary Mendelian Randomisation

To investigate the potential functional relevance of SNPs with disorder-specific effects, we applied the SMR approach [12], integrating eQTL (SNP-gene expression association) and mQTL (SNP_DNA methylation association) to the results from the conditional analyses. eQTL data from brain tissue were derived from a meta-analysis of the GTEx study, the Common Mind Consortium (CMC) and the Religious Orders Study and Memory and Aging Project (ROSMAP). The details of the meta-analysis have been described elsewhere[17]. Using meta-analysis results across brain tissues and studies is justified owing to the high correlation in effect sizes between tissues[17]. Only genes with a cis-eQTL with $p_{eQTL} < 5 \times 10^{-8}$ were included in the analysis. Experiment-wide significance accounting for testing multiple SNPs across multiple traits was set at $p_{SMR} = 1.9 \times 10^{-06}$ and the threshold for no evidence of heterogeneity due to pleiotropy at $p_{HEIDI} > 0.01$. Individual-level genotypes from the ARIC data (n = 7,762 unrelated individuals) [11] were used to estimate LD for the HEIDI test.

To test for the effects of disorder-specific variants on DNA methylation, we used SMR to integrating trait association data with meta-analysed brain mQTL data set from Jaffe et al. (n = 526) ROSMAP (n = 486) and fetal brain mQTL data from Hannon et al [18]. Only probes with at least one cis-mQTL with $p < 5 \times 10^{-8}$ were included in the SMR analysis. Probes that passed the significance threshold of 1.56×10^{-7} and did not show evidence of heterogeneity as indicated by the HEIDI test were considered to be significant.

Cell-type specificity for disorders

To gain insight into the cell types that are important for each disorder, we evaluated whether genes associated with specific brain cell-types are enriched for association with

each of the disorders. Using data from single-cell sequencing experiments in mice, the cell-type specificity of each gene was calculated by comparing the expression of a gene in a given cell-type to that across all cell types [19]. MAGMA was used to calculate gene-based association statistics and to evaluate whether genes with high specificity in a given cell-type are enriched for association with a disorder. The enrichment analysis was performed for both unadjusted and conditional GWAS for all 5 disorders. To investigate whether there was a significant change in the cell-type enrichment after conditioning, MAGMA analysis was performed using the enrichment Z-scores from the unadjusted GWAS as covariates in the analysis and a conditional enrichment for all level 1 cell types analysed in Skene et al. [19] was estimated.

Code Availability

Scripts used to generate the results are available on request from the corresponding author

Results

Baseline statistics

After merging GWAS summary statistics for the five psychiatric disorders 5,275,400 autosomal SNPs remained (Table 1). The number of independent genome-wide significant SNPs annotated by FUMA [14] is much greater for SCZ (M =130) compared to the other disorders (M =16, 40, 11, 2 for BIP, MD, ADHD, AUT respectively) reflecting mostly sample size, but also genetic architecture, and population risk. Linkage disequilibrium score regression (LDSC) estimates of SNP-based heritability on the liability scale and genetic correlations were all significantly different from zero (Table 2). Genetic correlations were highest between SCZ and BIP ($r_g = 0.67$ (s.e. = 0.03)) and lowest between BIP and ADHD ($r_g = 0.15$ (s.e. = 0.04)). The LD-score regression intercept was significantly greater than zero for the majority of pairs of disorders reflecting sample overlap in the GWAS studies. The intercept was highest between ADHD and AUT due to substantial overlap in controls. See Supplementary Material for discussion of interpretation of results in the context of sample overlap.

The GSMR analyses highlights some asymmetries in the estimates of the causal effects of one disorder on another (Table 3). In particular, the estimated liability $\hat{b}_{xy}$ when considering MD as an exposure for each trait is higher than the estimates in the reverse direction. One explanation is that since MD is so common and is frequently comorbid with other disorders that MD samples include those diagnosed and undiagnosed with other disorders. However, if model assumptions are violated it may have greater impact when there is a large difference in lifetime risk between the pairs of disorder. However, countering this, we find a higher $\hat{b}_{xy}$ from AUT to ADHD than from ADHD to AUT, but the standard errors on estimates are much higher for these disorders. Interpretation of these $\hat{b}_{xy}$ estimates depends on the nature of the shared genetic contributions to psychiatric disorders that may reflect a complex mix of types of pleiotropy, where some sets shared of variants may have more correlated effect sizes than other sets of shared variants.

Changes in Genetic Correlation

The impact of the conditioning is demonstrated by the changes in the estimates of $\widehat{r_g}$ comparing original and conditional GWAS results. The $\widehat{r_g}$ between SCZ conditional on the other disorders (denoted SCZ$_{cond}$) and SCZ remained high at 0.93, while between SCZ$_{cond}$ and BIP it was much reduced (from 0.67 prior to conditioning to 0.36, after conditioning). It is noted that b_{zy} is eliminated in the conditional analysis only if the SNP effect is mediated by trait x. Therefore, there is remaining genetic correlation because of pleiotropic SNP effects. A similar pattern of changes in genetic correlation with other traits was seen for the analyses with the other disorders as the outcome variable (Supplementary Table 1).

mtCOJO genome-wide significant SNP results

As expected because of pleiotropy between disorders, conditional analysis leads to a reduction in the mean test statistic across all SNPs in the genome and hence the number of independent SNPs reaching the significance threshold (5×10^{-8}) is reduced (Table 1). For each disorder, we present results for all independent SNPs significant in the unadjusted analysis or the conditional analysis (Supplementary Table 2). GWS SNPs that are more significantly associated in the conditional analysis than the unadjusted analysis are shown in Table 4. A larger conditional effect size suggests that these variants are disorder-specific or have heterogeneous effects across disorders.

Given that SCZ is the disorder with the largest number of significant SNPs and for which the power to detect changes in effects is largest, we focus mostly on the results from the SCZ conditional analysis. Of the 130 SNPs from the unadjusted SCZ GWAS, five were more significant after adjusting for the other disorders (all of which had opposite direction of effects for BIP – Supplementary Table 2) and a further eight had a larger estimated effect size after conditioning. Forest plots for the four most significant SCZ SNPs from the conditional analysis (two of which were associated $p < 5 \times 10^{-8}$ in the unadjusted analysis) are shown in Figure 1.

For all disorders except for AUT, a number of SNPs surpass the significance threshold that were not significant in the original GWAS. For schizophrenia, ten SNPs that were significant in the conditional analysis and not in the original GWAS (Table 4). All 10 SNPs have opposite effects for BIP, so that the allele that predisposes to SCZ is in the protective direction for BIP. Although these opposite effects could be due to ascertainment, among them are variants in or near genes with annotated biological functions that are potentially relevant for SCZ. For instance a SNP that was significant in the conditional analysis (rs2973038 – $p_{adj} = 1.28 \times 10^{-08}$; $p_{scz} = 1.72 \times 10^{-06}$) is located in the Glial Cell Derived Neurotrophic Factor (*GDNF*), a gene that encodes a protein that enhances the survival of midbrain dopaminergic neurons [20], and is expressed during development [21].

All SNPs that were associated with BIP at $p < 5 \times 10^{-8}$ in the original GWAS were less significant in the conditional analysis, showing evidence that they have some pleiotropic effect across disorders. Notably, this includes genes involved in calcium signalling, dopaminergic signalling and synaptic plasticity, indicating these processes may be important across psychiatric disorders. Three SNPs that were not significant in the BIP GWAS were

significant in the conditional analysis (Table 4, Supplementary Table 2, Supplementary Figure 2).

For each of the remaining disorders (MD, ADHD and AUT), we found that a small proportion of the existing significant SNPs had larger conditional effect sizes and one MD SNP and two ADHD SNPs that were not significant in the original GWAS became significant after conditioning (Table 4, Supplementary Table 2). However the difference in effect size after conditioning is not statistically significant for these SNPs, due to low statistical power (Supplementary Table 2). Forest plots for significant SNPs that had increased conditional effect sizes are shown in Supplementary Figures 3–5

SMR analysis

Changes in the expression of 9 genes were significantly associated with the 5 disorders (0 for BIP, 5 for SCZ, 3 for MD and 1 for ADHD, 0 for AUT) after conditioning and removal of genes in the MHC (Supplementary Tables 3–4), and a total of 72 DNA methylation sites (2 for BIP, 18 for SCZ, 37 for MD, 8 for ADHD, and 6 for AUT) were significantly associated with the 5 conditional traits (Supplementary Table 3–4).

Significant SMR results for gene expression where the associated SNP is more significant in the conditional analysis are presented in Supplementary Table 3. Three out of 5 significant SMR associations for SCZ were with SNPs where the conditional significance was greater than in the unadjusted analysis. One SNP - rs3759384 – is associated with decreased expression of *VPS29* in the brain and significantly increased risk for SCZ in the unadjusted analysis and has a larger conditional effect size (Supplementary Figure 6), indicating that *VPS29* may be linked to the development of SCZ and not other disorders. The VPS29 protein is a component of the retromer complex which prevents the degradation of certain proteins including signalling receptors, ion channels and small molecule transporters. The complex is essential for maintenance of neurons and has been implicated in the etiology of a number of neurodegenerative disorders [22].

One of the three associations for MD was with a SNP (rs7732179) with greater significance in the conditional analysis. The same variant shows evidence of association with SCZ but with opposite directions of effect ($b_{SCZ} = -0.045$; $p_{SCZ} = 1.7 \times 10^{-6}$ and $b_{BIP} = -0.029$; $p_{BIP} = 0.027$). The A allele confers risk to MDD but is protective for SCZ and BIP (Supplementary Figure 7). The SNP is associated with expression of *PCDHA7* in the brain. This gene encodes a member of the protocadherin family of genes located together on chromosome 5. A significant association was also found in this region in the DNA methylation analysis of MD. Little is known about the exact function of these genes, however they are concentrated at the synaptic junction suggesting a key role in neuronal signalling [23].

Out of 72 significant DNA methylation sites, 34 were associated with SNPs with higher significance in the conditional analyses (1 for BIP, 3 for SCZ, 21 for MD, 4 for ADHD and 5 for AUT) (Supplementary Table 3). It is noteworthy that one variant (rs2064853) was significantly associated with both SCZ and MD and DNA methylation near the *CSE1L* gene, but with opposite alleles increasing risk to each disorder (Supplementary Figure 8).

We investigated whether genes identified in the gene expression SMR or that are the closest gene to a significant methylation site are the primary target for FDA-approved drugs. We identified two genes that are targeted by medications. The serotonin receptor gene *HTR1D* which was identified in the DNA methylation analysis for MD is the primary target of the migraine drug naratriptan. Individuals with migraine are at 2–4 fold higher risk of developing depression and these results may suggest that triptans, used to treat migraines, could also be effective for MD.

The second drug target identified is with *MPL* and ADHD. This gene is targeted by romiplostim, an orphan drug developed for treatment of chronic idiopathic thrombocytopenic purpura.

MAGMA gene-set analyses

We conducted MAGMA gene-set analysis in FUMA to identify pathways and gene-sets that are enriched for association with the disorders after conditional analyses and to identify which sets become more or less significant after conditioning. Results for each disorder are presented in Supplementary Table 5. After conservative Bonferroni correction for the number of gene-sets tested for each disorder, three gene sets were significant - two for SCZ conditional analysis and one for AUT. For SCZ, the two significant sets were go:establishment of localization in the cell and GO:Dendrite, of which establishment of localization had a more significant p-value in the conditional analysis (Supplementary Table 5). For AUT, the gene-set GO:Dendrite_morphogenesis was significant after multiple testing and had a more significant p-value in the conditional analysis, potentially implicating genes expressed in dendrites in autism-specific pathology.

Cell-type specificity for disorders

The results from the cell-type enrichment analyses of raw and conditional analyses are shown in Figure 2. Consistent with previous results, the original SCZ results were enriched in medium spiny neurons (MSNs), pyramidal CA1 cells, pyramidal SS1 cells, interneurons and serotonergic neurons (Supplementary Table 6). All of these cell types also show some evidence of association with BIP and to a lesser extent MD, consistent with the genetic correlation between disorders and hence show reduced enrichment in the SCZ conditional analysis. All enriched cell-types for SCZ remained significant after conditioning except for serotonergic neurons, indicating that genes specific to this cell-type may increase risk to all five disorders. Enrichment in interneurons was found for SCZ, BIP and MDD indicating their potential importance across all 3 disorders. After conditioning, this cell-type was still significantly enriched in SCZ and MDD, but not BIP. This may reflect that the sample size of the BIP analysis is smaller than for SCZ and MDD.

Discussion

Our goal was to identify genetic variants that show disorder specific association by conducting a summary statistics based GWAS analysis for each of five psychiatric disorders conditioning on GWAS results from the other disorders. As expected, given the high degree of pleiotropy across disorders, compared to original GWAS results the number of SNPs

associated at the threshold of genome-wide significance is very much reduced for each conditional GWAS. We utilise mtCOJO as a method that uses summary statistics to quickly screen for SNP associations. Functional annotation can help prioritise the associations of most interest. It will be important to understand why a variant increases risk only to that disorder and not to others.

By integrating conditional GWAS results with SNP-gene expression and SNP-methylation results, we identify decreased expression of *VPS29* as a potential biological mechanism underlying schizophrenia. The variant that increases risk to SCZ and is associated with decreased expression of VPS29 in brain tissue shows no evidence for association with other psychiatric disorders. The retromer complex, of which is VPS29 is a subunit, is highly conserved across eukaryotes. The complex plays a role in the recycling, delivery and degradation of proteins in the cell and is crucial in the maintenance of cell homeostasis[24]. Rare exonic mutations in members of the complex have been associated with Parkinson's Disease and post-mortem studies have revealed decreased expression of all members of the complex in the brains of patients with Parkinson's and Alzheimer's disease. The expression of all three members of the complex is linked such that decreasing expression of one leads to decreased expression of all of them. Knocking down VPS35 using siRNA leads to elevated generation of amyloid-beta and reduced synaptic transmission[25, 26]. There is therefore considerable interest in identifying pharmacological agents that prevent the degradation of the retromer complex as a therapeutic mechanism for neurodegenerative disease. Small molecule screens have identified potential therapeutic agents that have shown promise in vitro[27]. Our results provide that such compounds may of interest in targeting biological mechnisms specific to schizophrenia.

Furthermore, SNPs associated with decreased expression of *PCDH7* and decreased methylation near other members of the protocadherin gene family on chromosome 5 may increase risk of MD, but be protective for SCZ and BIP. The protocadherins are a large family of genes involved in cell-cell adhesion that are primarily expressed in the nervous system. They play a major role in the development of the nervous system and in regulating dendritic branching. The PCDHA7 gene is part of a complex cluster of protocadherin alpha genes in the same genomic locus. The expression of the different isoforms at the locus is controlled by upstream CpG sites. Owing to their functional role in nervous system development and their location in linkage peaks, the PCDHA genes have been investigated as candidate genes for bipolar and schizophrenia[28]. Moreover, an epigenetic study of concordant and discordant MZ twins for depression showed that affected twins had increased variation in methylation in the PCDHA region, highlighting instability in this region as a potential mechanism underlying depression[29]. Further studies of the role of the PCDHA gene cluster in psychopathology are warranted.

Methylation in the promoter of the CSE1L gene, whose encoded protein influences cellular proliferation and has been linked to progression of a number of cancers, shows evidence of increasing risk to SCZ but being protective for MD.

Consistent with the large degree of pleiotropy between disorders, we found that most of the significant biological pathways for each disorder had reduced significance after

conditioning. Pathway analysis of conditional results identified a potential role for genes expressed in dendrites in both autism and schizophrenia. Likewise, for the cell-type enrichment analysis, there was a reduction in the enrichment for most cell types in each disorder after conditioning. For SCZ, the previously identified enrichments in pyramidal SS1 and CA1 cells as well as medium spiny neurons remained significant after conditioning, despite also showing evidence for enrichment in BIP. The largest change in enrichment was for serotonergic neurons, indicating that genes highly expressed there are important across all psychiatric disorders.

We provide an analysis framework for conditional cross-disorder analyses using summary statistics. Our study was motivated to improve on the SCZ case vs. BIP case analyses that utilised PGC cohorts for which both SCZ and BIP genotyped samples were available [30], but which necessarily excluded 28% of cases that could not be allocated into matched cohorts. They identified 5 SNPs associated at $p < 5 \times 10^{-8}$. We conducted an analysis of SCZ conditional on BIP and performed a lookup of those SNPs in the unadjusted and adjusted results. All but one (rs200005157) of their associated SNPs were matched directly or to an LD proxy (Supplementary Table 7). All show increased statistical significance in the conditional analysis. We identified more disorder-specific SNPs (10 specific to SCZ) consistent with the larger sample sizes afforded from using summary statistics, highlighting that mtCOJO is an efficient method for screening for disorder-specific SNPs for two or more related disorders. An in depth discussion of the mtCOJO method is given in the Supplementary Material.

Limitations

There are a number of limitations to our analyses that should be considered. Although methods that utilise summary statistics have several advantages, they also depend upon the summary statistics being generated accurately. In this instance, all studies have gone through the same quality control and analysis pipeline meaning that systematic differences between studies are unlikely. It is not clear how misdiagnosis of cases would impact upon the results.

There are also substantial differences in sample size between the GWAS of different disorders, with SCZ and MD having a larger sample size than the other disorders, which may disproportionately influence the results. This is shown by most of the significant differences in effect sizes between the raw and conditional results being for SCZ. The disorders that have the most genome-wide significant SNPs will also have the most accurate estimates of their effects on the disorders. As sample sizes increase for some of the other disorders, the results for those disorders will become more accurate.

In order to reduce the burden of multiple testing in the SMR analysis, we only included SNPs that are associated at the genome-wide significant level with gene expression or methylation in cis. Relaxing the statistical threshold for inclusion may have identified more SNPs with effects on gene expression in brain with the trade-off of increasing the experiment-wide significance level.

Conclusion

In conclusion, our results suggest that mtCOJO is an efficient method for identifying variants with disorder-specific effects and they represent a small fraction of variants identified for each disorder to date, reflecting the high degree of pleiotropy between disorders. Nonetheless, we identify several loci that have evidence of being disorder-specific. Further research in human studies should focus on whether the disorder-specific variants associate with specific symptoms in mixed clinical populations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Conflict of Interest Statement

P.F.S. is on the advisory committee at Lundbeck, is a Scientific Advisory Board member at Pfizer and has received speaker reimbursement and grant funding from Roche. J.H.-L. is a Scientific Advisor at Cartana and has received grant funding from Roche.

References

1. Cross-Disorder Group of the Psychiatric Genomics Consortium, Lee SH, Ripke S, Neale BM, Faraone SV, Purcell SMet al.Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nature genetics2013; 45(9): 984–994. [PubMed: 23933821]

2. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PRet al.An atlas of genetic correlations across human diseases and traits. Nature genetics2015; 47(11): 1236–1241. [PubMed: 26414676]

3. Grove J, Ripke S, Als TD, Mattheisen M, Walters R, Won Het al.Common risk variants identified in autism spectrum disorder. bioRxiv2017.

4. Cross-Disorder Group of the Psychiatric Genomics C. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. Lancet2013; 381(9875): 1371–1379. [PubMed: 23453885]

5. Turley P, Walters RK, Maghzian O, Okbay A, Lee JJ, Fontana MAet al.Multi-trait analysis of genome-wide association summary statistics using MTAG. Nature genetics2018.

6. Maier R, Moser G, Chen GB, Ripke S, Cross-Disorder Working Group of the Psychiatric Genomics C, Coryell Wet al.Joint analysis of psychiatric disorders increases accuracy of risk prediction for schizophrenia, bipolar disorder, and major depressive disorder. American journal of human genetics2015; 96(2): 283–294. [PubMed: 25640677]

7. Scott J, Leboyer M, Hickie I, Berk M, Kapczinski F, Frank Eet al.Clinical staging in psychiatry: a cross-cutting model of diagnosis with heuristic and practical value. The British journal of psychiatry : the journal of mental science2013; 202(4): 243–245. [PubMed: 23549937]

8. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui Aet al.Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nature genetics2018; 50(5): 668–681. [PubMed: 29700475]

9. Nieuwboer HA, Pool R, Dolan CV, Boomsma DI, Nivard MG. GWIS: Genome-Wide Inferred Statistics for Functions of Multiple Phenotypes. American journal of human genetics2016; 99(4): 917–927. [PubMed: 27616482]

10. Zhu Z, Zheng Z, Zhang F, Wu Y, Trzaskowski M, Maier Ret al.Causal associations between risk factors and common diseases inferred from GWAS summary data. Nat Commun2018; 9(1): 224. [PubMed: 29335400]

11. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. American journal of epidemiology1989; 129(4): 687–702. [PubMed: 2646917]

12. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JEet al.Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nature genetics2016; 48(5): 481–487. [PubMed: 27019110]

13. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. American journal of human genetics2011; 88(1): 76–82. [PubMed: 21167468]

14. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun2017; 8(1): 1826. [PubMed: 29184056]

15. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput Biol2015; 11(4): e1004219. [PubMed: 25885710]

16. Zheng J, Erzurumluoglu AM, Elsworth BL, Kemp JP, Howe L, Haycock PCet al.LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. Bioinformatics2017; 33(2): 272–279. [PubMed: 27663502]

17. Qi T, Wu Y, Zeng J, Zhang F, Xue A, Jiang Let al.Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. Nat Commun2018; 9(1): 2282. [PubMed: 29891976]

18. Hannon E, Dempster E, Viana J, Burrage J, Smith AR, Macdonald Ret al.An integrated genetic-epigenetic analysis of schizophrenia: evidence for co-localization of genetic associations and differential DNA methylation. Genome Biol2016; 17(1): 176. [PubMed: 27572077]

19. Skene NG, Bryois J, Bakken TE, Breen G, Crowley JJ, Gaspar HAet al.Genetic identification of brain cell types underlying schizophrenia. Nature genetics2018; 50(6): 825–833. [PubMed: 29785013]

20. Lin LF, Doherty DH, Lile JD, Bektesh S, Collins F. GDNF: a glial cell line-derived neurotrophic factor for midbrain dopaminergic neurons. Science1993; 260(5111): 1130–1132. [PubMed: 8493557]

21. Stromberg I, Bjorklund L, Johansson M, Tomac A, Collins F, Olson Let al.Glial cell line-derived neurotrophic factor is expressed in the developing but not adult striatum and stimulates developing dopamine neurons in vivo. Exp Neurol1993; 124(2): 401–412. [PubMed: 7904571]

22. Wang S, Bellen HJ. The retromer complex in development and disease. Development2015; 142(14): 2392–2396. [PubMed: 26199408]

23. Yagi T, Takeichi M. Cadherin superfamily genes: functions, genomic organization, and neurologic diversity. Genes Dev2000; 14(10): 1169–1180. [PubMed: 10817752]

24. Vagnozzi AN, Pratico D. Endosomal sorting and trafficking, the retromer complex and neurodegeneration. Molecular psychiatry2019; 24(6): 857–868. [PubMed: 30120416]

25. Bhalla A, Vetanovetz CP, Morel E, Chamoun Z, Di Paolo G, Small SA. The location and trafficking routes of the neuronal retromer and its role in amyloid precursor protein transport. Neurobiol Dis2012; 47(1): 126–134. [PubMed: 22516235]

26. Zhang H, Huang T, Hong Y, Yang W, Zhang X, Luo Het al.The Retromer Complex and Sorting Nexins in Neurodegenerative Diseases. Front Aging Neurosci2018; 10: 79. [PubMed: 29632483]

27. Mecozzi VJ, Berman DE, Simoes S, Vetanovetz C, Awal MR, Patel VMet al.Pharmacological chaperones stabilize retromer to limit APP processing. Nat Chem Biol2014; 10(6): 443–449. [PubMed: 24747528]

28. Pedrosa E, Stefanescu R, Margolis B, Petruolo O, Lo Y, Nolan Ket al.Analysis of protocadherin alpha gene enhancer polymorphism in bipolar disorder and schizophrenia. Schizophrenia research2008; 102(1–3): 210–219. [PubMed: 18508241]

29. Cordova-Palomera A, Fatjo-Vilas M, Gasto C, Navarro V, Krebs MO, Fananas L. Genome-wide methylation study on depression: differential methylation and variable methylation in monozygotic twins. Transl Psychiatry2015; 5: e557. [PubMed: 25918994]

30. Bipolar D, Schizophrenia Working Group of the Psychiatric Genomics Consortium. Electronic address drve, Bipolar D, Schizophrenia Working Group of the Psychiatric Genomics C. Genomic Dissection of Bipolar Disorder and Schizophrenia, Including 28 Subphenotypes. Cell2018; 173(7): 1705–1715 e1716. [PubMed: 29906448]

rs3764002 − mtCOJO p−value = 1.9 x 10E−12
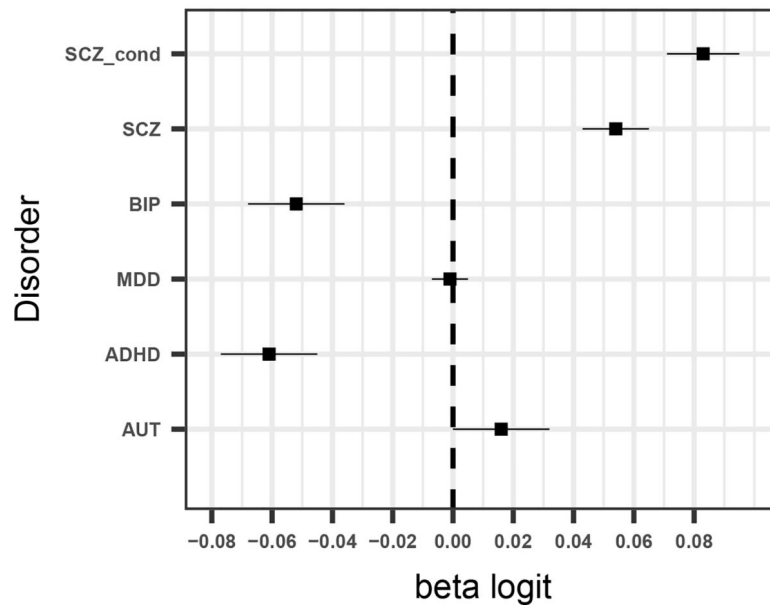
rs6701877 − mtCOJO p−value = 1.47 x 10E−11
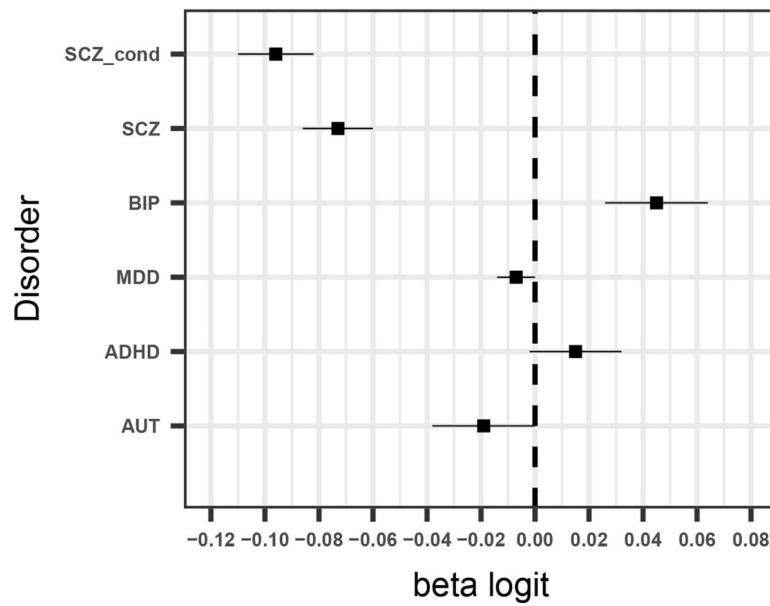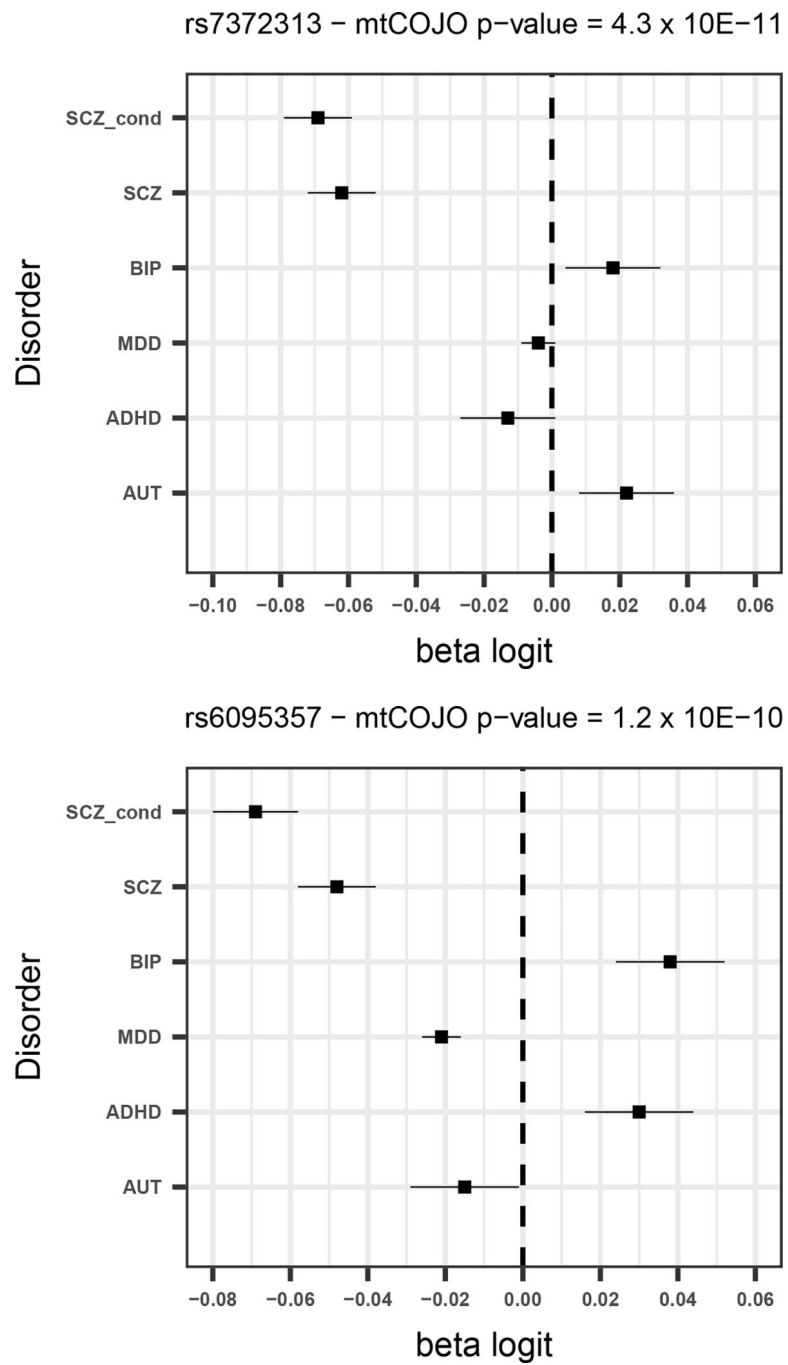
Figure 1 –.
Forest plots for the 4 most significant SNPs in SCZ mtCOJO analysis with larger conditional effect sizes
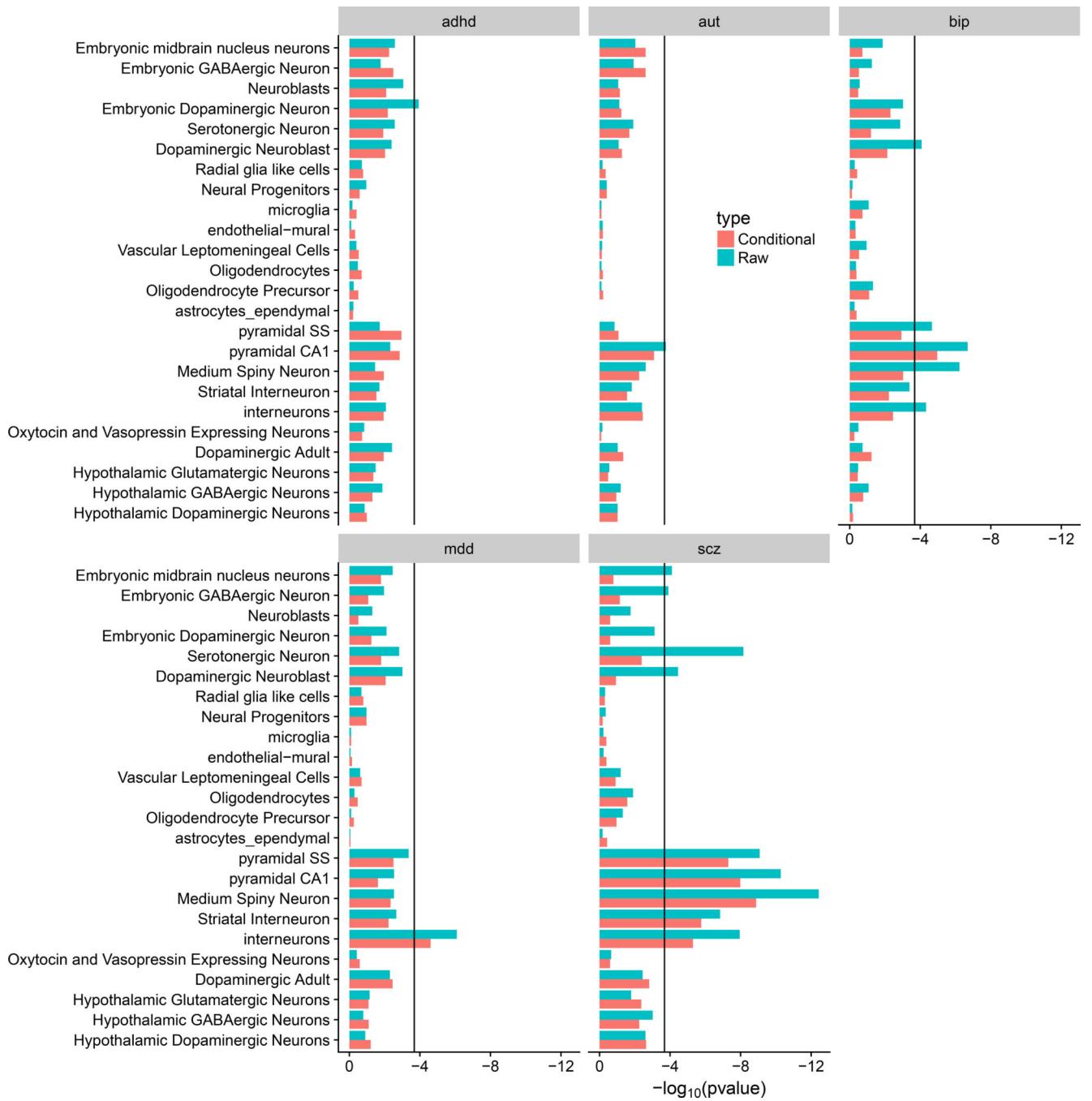
Figure 2.
Results from brain cell-type enrichment analyses of raw and conditional GWAS analyses

Table 1.

Summary of datasets used and results from conditional GWAS analysis

Disorder	Cases	Controls	No. SNPs in original study	No GWS SNPs in merged data	Study reference	No. GWS SNPs in Published Study	Assumed Lifetime Disease Risk	GWS loci in conditional GWAS	New GWS loci	GWS SNPs from unadjusted GWAS with larger conditional effect size
SCZ	40,675	64,643	5,471,613	130	Pardinas et al 2018	145	0.01	43	10	15
BIP	20,352	31,358	9,498,970	16	Stahl et al 2018 BioRxiv	19	0.01	4	3	0
MD	135,458	344,901	10,468,943	40	Wray et al 2018	44	0.15	15	1	5
ADHD	19,099	34,194	6,755,648	11	Demontis et al BioRxiv	12	0.05	5	2	3
AUT	18,381	27,969	7,539,669	2	Grove et al BioRxiv	3	0.01	1	0	1

Table 2.

Estimated SNP-based heritability on the liability scale, genetic correlation and LD-score intercepts estimated from LD-score regression

	SCZ	BIP	MDD	ADHD	AUT
SCZ	**0.23 (0.01)**	*0.21 (0.01)*	*0.03 (0.01)*	0.02 (0.01)	0.008 (0.01)
BIP	0.67 (0.02)	**0.19 (0.01)**	*0.05 (0.007)*	*0.03 (0.006)*	0.009 (0.008)
MD	0.36 (0.02)	0.35 (0.02)	**0.08 (0.004)**	*0.10 (0.008)*	*0.09 (0.008)*
ADHD	0.18 (0.03)	0.15 (0.04)	0.43 (0.03)	**0.22 (0.01)**	*0.35 (0.008)*
AUT	0.23 (0.05)	0.15 (0.05)	0.43 (0.04)	0.36 (0.05)	**0.12 (0.01)**

LD-score SNP-based heritability on the liability scale and standard error reported on diagonal

r_g and standard error reported below the diagonal

Bivariate ldsc intercept reported above the diagonal. Value significantly greater than zero (in italics) quantify sample overlap

Table 3.

GSMR estimates of causal effect of each psychiatric disorder on the others with conversion to the log odds ratio and liability scales

Exposure	Outcome	N SNPs	bxy	bxy_se	bxy_liability	OR	bxy_pval
SCZ	BIP	111	0.417	0.019	0.417	3.06	5.0E-109
SCZ	MD	111	0.074	0.007	0.109	1.22	4.9E-26
SCZ	ADHD	111	0.054	0.019	0.066	1.16	5.2E-03
SCZ	AUT	111	0.144	0.019	0.144	1.47	2.9E-09
BIP	SCZ	16	0.498	0.039	0.498	3.82	1.6E-37
BIP	MD	16	0.091	0.016	0.134	1.28	2.0E-08
BIP	ADHD	16	0.028	0.043	0.034	1.08	5.2E-01
BIP	AUT	16	0.123	0.046	0.123	1.39	7.4E-03
MD	SCZ	40	0.414	0.059	0.281	2.13	2.7E-12
MD	BIP	40	0.600	0.068	0.408	2.97	1.1E-18
MD	ADHD	40	0.402	0.072	0.339	2.09	2.9E-08
MD	AUT	40	0.463	0.078	0.314	2.33	3.7E-11
ADHD	BIP	13	0.135	0.052	0.109	1.34	8.9E-03
ADHD	MD	13	0.086	0.019	0.102	1.21	9.1E-06
ADHD	SCZ	13	0.156	0.043	0.126	1.40	2.8E-04
ADHD	AUT	11	0.333	0.060	0.269	2.06	2.9E-08
AUT	SCZ	11	0.063	0.041	0.063	1.19	1.3E-01
AUT	BIP	11	0.053	0.057	0.053	1.15	2.9E-01
AUT	MD	11	0.011	0.021	0.016	1.03	5.9E-01
AUT	ADHD	11	0.413	0.062	0.512	3.03	3.60E-11

*
Estimates using autism as the exposure used instruments with p < 10E-06 due to lack of genome-wide significant SNPs for autism

Table 4.

Results for SNPs that were genome-wide significant in the conditional analysis and have larger estimated conditional effect sizes than in the original GWAS

Disorder	SNP	CHR	Position	A1	Adjusted beta	SE Adjusted beta	Unadjusted beta	SE Unadjusted beta	Adjusted p-value	Unadjusted p-value	nearestGene
SCZ	rs3764002	12	108618630	C	0.083	0.012	0.054	0.011	1.94E-12	6.05E-07	WSCD2
SCZ	rs6095357	20	47523865	A	-0.069	0.011	-0.048	0.010	1.17E-10	1.21E-06	ARFGEF2
SCZ	rs7790864	7	28478625	A	-0.062	0.011	-0.044	0.010	6.33E-09	7.18E-06	CREB5
SCZ	rs1054972	19	1852582	A	0.076	0.013	0.053	0.012	6.42E-09	1.32E-05	KLF16
SCZ	rs2867673	7	71752652	T	0.060	0.010	0.049	0.010	9.44E-09	4.11E-07	CALN1
SCZ	rs6564668	16	79457393	C	-0.060	0.010	-0.038	0.010	1.05E-08	7.94E-05	RP11-467I7.1
SCZ	rs11922765	3	95047279	G	-0.060	0.010	-0.044	0.010	1.22E-08	4.36E-06	RPS18P6
SCZ	rs2973038	5	37833781	C	0.066	0.012	0.051	0.011	1.28E-08	1.72E-06	GDNF
SCZ	rs10903945	10	363275	C	0.057	0.010	0.040	0.010	3.13E-08	3.30E-05	DIP2C
SCZ	rs10282935	8	38703797	A	0.058	0.011	0.041	0.010	3.97E-08	3.17E-05	TACC1
SCZ	rs6701877	1	174015259	G	-0.096	0.014	-0.073	0.013	1.47E-11	2.37E-08	RP11-160H22.3
SCZ	rs7372313	3	135872958	G	-0.069	0.010	-0.062	0.010	4.26E-11	1.54E-10	MSL2
SCZ	rs1765142	11	30378559	C	0.065	0.011	0.058	0.010	1.54E-09	1.13E-08	ARL14EP
SCZ	rs55646993	7	105017864	G	-0.062	0.010	-0.053	0.010	2.23E-09	3.83E-08	SRPK2
SCZ	rs150437760	14	59981768	A	0.131	0.024	0.121	0.022	3.71E-08	4.58E-08	CCDC175
BIP	rs12554512	9	23352293	T	-0.083	0.014	-0.066	0.014	1.55E-09	1.28E-06	ELAVL2
BIP	rs6891181	5	80849101	T	-0.081	0.014	-0.075	0.014	1.49E-08	1.27E-07	SSBP2
BIP	rs12268910	10	111878510	T	-0.097	0.018	-0.091	0.018	3.29E-08	2.73E-07	ADD3
MD	rs11697370	20	47731767	T	-0.031	0.005	-0.023	0.005	3.31E-09	3.53E-06	STAU1
MD	rs27732	5	87992576	A	0.034	0.005	0.031	0.005	1.22E-11	1.87E-10	MEF2C
MD	rs1806153	11	31850105	T	0.037	0.006	0.036	0.006	8.78E-10	1.18E-09	RCN1
MD	rs1354115	9	2983774	A	0.029	0.005	0.028	0.005	1.72E-08	2.37E-08	CARM1P1
MD	rs301799	1	8489302	T	-0.028	0.005	-0.026	0.005	2.49E-08	4.68E-08	RERE
ADHD	rs78648104	6	50683009	T	0.136	0.023	0.124	0.025	4.31E-09	3.60E-07	TFAP2D
ADHD	rs2244336	10	8831827	C	0.071	0.013	0.069	0.014	3.81E-08	3.67E-07	ENSG00000270234

Disorder	SNP	CHR	Position	A1	Adjusted beta	SE Adjusted beta	Unadjusted beta	SE Unadjusted beta	Adjusted p-value	Unadjusted p-value	nearestGene
ADHD	rs12410444	1	44188719	A	0.107	0.014	0.106	0.015	4.23E-15	3.85E-13	ST3GAL3
ADHD	rs13023832	2	215219808	A	0.133	0.020	0.117	0.021	1.23E-11	1.62E-08	SPAG16
ADHD	rs281320	15	47769424	T	−0.080	0.013	−0.074	0.013	1.84E-10	3.14E-08	SEMA6D
AUT	rs10099100	8	10576775	C	0.084	0.014	0.084	0.015	1.20E-09	1.07E-08	SOX7