Genomic dissection of bipolar disorder and schizophrenia including 28 subphenotypes

Primary Author List
Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium

Secondary Author List
Summary

Schizophrenia and bipolar disorder are two distinct diagnoses that share symptomology, understanding the genetic factors contributing to the shared and disorder-specific symptoms will be crucial for improving diagnosis and treatment. In genetic data consisting of 53,555 cases (20,129 BD, 33,426 SCZ) and 54,065 controls, we identified 114 genome-wide significant loci implicating synaptic and neuronal pathways shared between disorders. Comparing SCZ to BD (23,585 SCZ, 15,270 BD) identified four genomic regions including one with disorder-independent causal variants and potassium ion response genes as contributing to differences in biology between the disorders. Polygenic risk scores (PRS) identified several significant correlations within case-only phenotypes including SCZ PRS with psychotic features and age of onset in BD. For the first time, we discover specific loci that distinguish between BD and SCZ and identify polygenic components underlying multiple symptom dimensions. These results point to the utility of genetics to inform symptomology and potentially treatment.

Introduction

Bipolar disorder (BD) and schizophrenia (SCZ) are severe psychiatric disorders and among the leading causes of disability worldwide(Whiteford et al., 2013). Both disorders have significant genetic components with heritability estimates ranging from 60-80%(Nöthen et al., 2010). Recent genetic and epidemiological studies have demonstrated substantial overlap between these two disorders with a genetic correlation from common variation near 0.6-0.7(Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013) and high relative risks (RR) among relatives of both BD and SCZ patients (RRs for parent/offspring: BD/BD: 6.4, BD/SCZ: 2.4;
SCZ/BD: 5.2, SCZ/SCZ: 9.9) (Lichtenstein et al., 2009). Despite shared genetics and symptomology, the current diagnostic systems (“Diagnostic and Statistical Manual of Mental Disorders | DSM Library,” n.d.) (“WHO | International Classification of Diseases,” n.d.) adhere to historical distinctions from the late 19th century and represent BD and SCZ as independent categorical entities differentiated on the basis of their clinical presentation, with BD characterized by predominant mood symptoms, mood-congruent delusions and an episodic disease course and SCZ considered a prototypical psychotic disorder. Identifying genetic components contributing to both disorders provides insight into the biology underlying the shared symptoms of the disorders.

While the shared genetic component is substantial, studies to date have also implicated genetic architecture differences between these two disorders (Curtis et al., 2011; Ruderfer et al., 2014). A polygenic risk score created from a case only SCZ vs BD genome-wide association study (GWAS) significantly correlated with SCZ or BD diagnosis in an independent sample (Ruderfer et al., 2014), providing the first evidence that differences between the disorders also have a genetic basis. An enrichment of rare, moderate to highly penetrant copy number variants (CNVs) and de novo CNVs are seen in SCZ patients (CNV and Schizophrenia Working Groups of the Psychiatric Genomics Consortium, 2017; Gulsuner and McClellan, 2015; Kirov et al., 2012; Stone et al., 2008; Szatkiewicz et al., 2014), while, the involvement of CNVs in BD is less clear (Green et al., 2016). Although the role of de novo single nucleotide variants in BD and SCZ have been investigated in only a handful of studies, enrichment in pathways associated with the postsynaptic density has been reported for SCZ, but not BD (Fromer et al., 2014; Kataoka et al., 2016). Identifying disorder-specific variants and quantifying the contribution of genetic variation to specific symptom dimensions remain important open questions. These genetic differences will
facilitate an understanding of the dimensions of the disorders instead of the dichotomous diagnosis. For example, we have shown that SCZ patients with greater manic symptoms have higher polygenic risk for BD (Ruderfer et al., 2014). These findings demonstrate shared genetic underpinnings for symptoms across disorders and may enable us to characterize patients by genetic liability to symptom dimensions thereby informing disease course and treatment.

Here, we utilize large collections of genotyped samples for BD and SCZ along with measures identifying 28 subphenotypes to address three questions: 1) Are there specific variants, genes or pathways that are either shared by, or differentiate BD and SCZ? 2) Are the shared symptoms between these disorders driven by the same underlying genetic profiles? and 3) Can we demonstrate independent genetic signatures for subphenotypes within these disorders?

**Results**

**Shared genetic contribution to BD and SCZ**

We performed association analysis of BD and SCZ combined into a single phenotype, totaling 53,555 cases (20,129 BD, 33,426 SCZ) and 54,065 controls on 15.5 million SNP allele dosages imputed from 1000 genomes phase 3 (The 1000 Genomes Project Consortium, 2015). Logistic regression was performed controlling for 13 principal components of ancestry, study sites and genotyping platform. We identified 11,231 SNPs with p-value below our genome-wide significance (GWS) threshold of $5 \times 10^{-8}$. After grouping SNPs in linkage disequilibrium with each other ($r^2 > 0.2$), 114 genomic risk loci remained. For the most significant variant in each of the 114 GWS loci, we performed conditional analysis with any GWS hit within 1Mb of the extent of the locus from the previously performed single disease GWAS of SCZ (Schizophrenia...
Working Group of the Psychiatric Genomics Consortium, 2014) and BD (Stahl et al., 2017) and identified 32 loci that were independently significant defined strictly as no single disease locus within 1Mb or a GWS p-value after conditional analysis (Supplementary Table 1). We further performed gene-set based tests using MAGMA (Leeuw et al., 2015) across 10,891 curated pathways (Watanabe et al., 2017) and identified 8 pathways surpassing Bonferroni correction (p < 4.6x10^-6) with all but one pathway implicating synaptic and neuronal biology (Supplementary Table 2a). Establishing independent controls (see Methods) allowed us to perform disorder-specific GWAS in 20,129 BD cases vs 21,524 BD controls and 33,426 SCZ cases and 32,541 SCZ controls. Using these results, we compared effect sizes of these 114 loci across each disorder independently (Figure 1a) showing the subsets of variants that had larger effects in SCZ vs BD and vice versa.

**Differentiating genetic contribution to BD and SCZ**

To identify loci with divergent effects on BD and SCZ, we performed an association analysis comparing 23,585 SCZ cases with 15,270 BD cases matched for shared ancestry and genotyping platform (see Methods, Figure 1b. Supplementary Figures 1-2, Table 1). Two genome-wide significant loci were identified, the most significant of which was rs56355601 located on chromosome 1 at position 173,811,455 within an intron of DARS2. The second most significant locus was rs200005157, a four base-pair insertion/deletion, on chromosome 20 at position 47638976 in an intron of ARFGEF2. For both variants, the minor allele frequency was higher in BD cases than SCZ cases and disease-specific GWAS showed opposite directions of effect when compared to controls. We sought to identify additional disease-specific loci by comprehensively incorporating expression information with association results to perform fine-mapping and
identify novel variants (Gamazon et al., 2015; Giambartolomei et al., 2014; Gusev et al., 2016; He et al., 2013). Here, we applied the summary-data-based Mendelian randomization (SMR) method (Zhu et al., 2016) (see Methods) utilizing the cis-QTLs derived from peripheral blood (Westra et al., 2013), human dorsolateral prefrontal cortex (DLPFC) (Fromer et al., 2016) from the Common Mind Consortium and 11 brain regions from the GTEx consortium (Consortium, 2015). We identified one SNP-probe combination that surpassed the threshold for genome-wide significance in blood but was also the most significant finding in brain. We found that SNP rs4793172 in gene DCAKD is associated with SCZ vs BD analysis ($p_{GWAS} = 2.8 \times 10^{-6}$) and is an eQTL for probe ILMN 1811648 ($p_{eQTL} = 2.9 \times 10^{-168}$), resulting in $p_{SMR} = 4.1 \times 10^{-6}$ in blood ($p_{eQTL} = 2.9 \times 10^{-25}, p_{SMR} = 2.0 \times 10^{-5}$ in DLPFC, and $p_{eQTL} = 4.6 \times 10^{-15}, p_{SMR} = 6.0 \times 10^{-5}$ in GTEx cerebellar hemisphere) (Supplementary Table 3, Supplementary Figure 3) and shows no evidence of heterogeneity ($p_{HET} = 0.66$) which implies only a single causal variant in the locus.

In an effort to prioritize genes for the two GWS loci from the GWAS, we performed fine-mapping (Benner et al., 2016) using an LD map derived from a majority of the control samples. We then performed SMR on each of the variants with causal probability greater than 1% using all eQTLs from the CommonMind Consortium DLPFC reference. All the most likely causal variants were shown to most significantly regulate the same gene suggesting CSE1L is the most likely relevant gene on chromosome 20 (rs200005157: causal probability=0.21, $p_{GWAS}=2.4 \times 10^{-8}$, $p_{eQTL} = 3 \times 10^{-8}, p_{SMR}=8.5 \times 10^{-5}, p_{HET}=0.34$, Supplementary Table 4). For the locus on chromosome 1, SLC9C2 is the most significantly regulated gene. However, a highly significant heterogeneity test indicates a complex genetic architecture making it difficult to infer a causal role for the associated SNP (Supplementary Table 5). Therefore, DARS2 presents as the most likely relevant
gene on chromosome 1 (rs56355601: $p_{\text{GWAS}}=5.6 \times 10^{-9}$, causal probability=$0.079$, $p_{\text{eQTL}}=7.4 \times 10^{-13}$, $p_{\text{SMR}}=6.17 \times 10^{-6}$, $p_{\text{HET}}=0.03$). We note however, that in both cases there are less associated variants that are stronger eQTLs for these genes complicating a straightforward causal interpretation. Finally, using the same gene-set test used for the combined analysis GO biological process “response to potassium ion” ($p=1.6 \times 10^{-6}$) was the only pathway surpassing our Bonferroni corrected significance threshold (Supplementary Table 2b).

Regional joint association

We expanded our efforts to identify disorder-specific genomic regions by jointly analyzing independent GWAS results from BD and SCZ(Pickrell et al., 2016). The genome was split into 1,703 previously defined approximately LD independent regions(Berisa and Pickrell, 2015). Thirteen percent, or 223 regions, had a posterior probability greater than 0.5 of having a causal variant for at least one disorder. Of these, 132 best fit the model of a shared causal variant influencing both BD and SCZ, 88 were most likely specific to SCZ, 3 demonstrated evidence of two independent variants (with one impacting each of the two disorders) and none were BD-specific. Of note, this approach calculates a prior probability that any given region is disease-specific and from these data the probability of having a BD specific region was 0.1% compared to 15% for SCZ, likely a result of increased power from the larger SCZ sample size and/or a difference in genetic architecture between these disorders.

The 114 GWS SNPs from the combined BD and SCZ GWAS localized into 99 independent regions (13 regions had multiple GWS SNPs), of which 78 (79%) were shared with a posterior probability of greater than 0.5. Sixty regions had at least one GWS SNP in the independent SCZ GWAS, of which 30 (50%) are shared and 8 regions contained a GWS SNP in the independent
BD GWAS, of which 6 (75%) are shared using the same definition. For the three regions showing evidence for independent variants, two had highly non-overlapping association signals in the same region stemming from independent variants. The third, on chromosome 19 presented a different scenario where association signals were overlapping (Figure 2). The most significant variant in BD was rs111444407 (chr19:19358207, p = 8.67x10^{-10}) and for SCZ was rs2315283 (chr19:19480575, p=4.41x10^{-7}). After conditioning on the most significant variant in the other disorder, the association signals of the most significant variant in BD and SCZ were largely unchanged (BD rs111444407 =1.3x10^{-9}, SCZ rs2315283 p=6.7x10^{-5}). We further calculated the probability of each variant in the region being causal for both BD and SCZ(Benner et al., 2016) and found no correlation (r= -0.00016). The most significant variants had the highest posterior probability of being causal (SCZ: rs2315283, prob = 0.02, BD: rs111444407, prob = 0.16). Both variants most significantly regulate the expression of GATAD2A in brain(Fromer et al., 2016) but in opposite directions (rs111444407 p_{eQTL} = 6x10^{-15}, beta = 0.105; rs2315283 p_{eQTL} = 1.5x10^{-28}, beta = -0.11).

Additional work calculating heritability estimates among the same set of 1,703 regions found comparable h^2 estimates across the genome except for the major histocompatibility complex on chromosome 6 and a single region on chromosome 10 (see Supplement).

**Regional SNP-heritability estimation**

Across the genome, regional SNP-heritabilities (h^2_{snp}) were estimated separately for SCZ and BD(Shi et al., 2016) and were found to be moderately correlated (r=0.25). We next defined risk regions as those containing the most associated SNP for each GWS locus. In total, there were 101 SCZ risk regions from the 105 autosomal GWS loci reported previously(Schizophrenia...
Working Group of the Psychiatric Genomics Consortium, 2014) and 29 BD risk regions from 30
GWS loci reported in a companion paper (Stahl et al., 2017). Ten regions were risk regions for
both BD and SCZ comprising 33% of BD risk regions and 10% of SCZ risk regions. We further
stratified regional \( h^2_{\text{snp}} \) by whether a region was a risk region in one disorder, none or both
(Supplementary Figure 4). Since the discovery data for the regions overlapped with the data used
for the heritability estimation, we expected within-disorder analyses to show significant results.
In risk regions specific to SCZ (n=91) there was a significant increase in regional \( h^2_{\text{snp}} \) in SCZ, as
expected \((p = 1.1\times10^{-22})\), but also in BD \((p = 1.2\times10^{-6})\). In risk regions specific to BD (n=19),
significantly increased regional \( h^2_{\text{snp}} \) was observed in BD, as expected \((p = 0.0007)\), but not in
SCZ \((p = 0.89)\). Risk regions shared by both disorders had significantly higher \( h^2_{\text{snp}} \) in both
disorders, as expected \((BD \ p = 5.3\times10^{-5}, \ SCZ \ p = 0.006)\), compared to non-risk regions.
However, we observed a significant increase in BD \( h^2_{\text{snp}} \) in shared risk regions compared to BD
risk regions \((BD \ p = 0.003)\) but not SCZ \( h^2_{\text{snp}} \) for shared risk regions compared to SCZ risk
regions \((p = 0.62)\). Using a less stringent p-value threshold for defining risk regions \((p < 5\times10^{-6})\),
thereby substantially increasing the number of regions, resulted in similar results. Seven regions
contributed to substantially higher \( h^2_{\text{snp}} \) in SCZ compared to BD but no region showed the
inverse pattern. Of these regions, all but one was in the major histocompatibility region (MHC),
the sole novel region was chr10:104380410-106695047 with regional \( h^2_{\text{snp}} = 0.0019 \) in SCZ and
\( h^2_{\text{snp}} = 0.00063 \) in BD.

**Polygenic dissection of subphenotypes**

Subphenotypes were collected for a subset of patients with both BD and SCZ (see Methods). For
SCZ, we had clinical quantitative measurements of manic, depressive, positive and negative
symptoms generated from factor analysis of multiple instruments as described previously (Ruderfer et al., 2014) but in larger sample sizes (n=6908, 6907, 8259, 8355 respectively). For BD, 24 subphenotypes were collected among nearly 13,000 cases in distinct categories including comorbidities, clinical information such as rapid cycling and psychotic features as well as additional disease course data such as age of onset and number of hospitalizations. For each BD and SCZ patient, we calculated a polygenic risk score (PRS) using all SNPs, from each of the four main GWAS analyses (BD+SCZ, BD, SCZ and SCZvsBD). We then used regression analysis including principal components and site to assess the relationship between each subphenotype and the 4 PRS. Specifically, we tested whether polygenic risk scores of BD+SCZ, BD, SCZ or SCZvsBD were correlated with each of these subphenotypes separately within BD and SCZ cases. When testing if the variance explained by the PRS was different from zero, we applied a significance cutoff of p < 0.0004 based on Bonferroni correction for 112 tests. In total, we identified 6 significant results after correction (Figure 3, Table 2).

A significant positive correlation existed between BD PRS and manic symptoms in SCZ cases as seen previously (Ruderfer et al., 2014) (p=2x10⁻⁵, t=4.26) and BD PRS and psychotic features in BD patients (p=5.3x10⁻⁵, t=4.04). A significant increase in SCZ PRS was seen for BD cases with versus without psychotic features (p=1.2x10⁻¹⁰, t=6.45) and patients with increased negative symptoms in SCZ patients (p=3.60x10⁻⁶, t=4.64). The BD+SCZ vs controls PRS was significantly associated with psychotic features in BD (p=7.9x10⁻¹³, t=7.17) and negative symptoms in SCZ (p=1.5x10⁻⁵, t=4.33). The next two most significant results which did not survive our conservative correction were both indicative of a more severe course in BD: increased BD+SCZ PRS with increased numbers of hospitalizations in BD cases (p=4.2x10⁻⁴,
t=3.53) and increased SCZ PRS with earlier onset of BD (p=7.9x10^{-4}, t=-3.36). We assessed the role of BD subtype on the correlation between SCZ PRS and psychotic features and identified a significant correlation when restricted to only BD type I cases indicating the result was not likely driven by BD patients with a schizoaffective subtype (BDI: 3,763 with psychosis, 2,629 without, p=1.55x10^{-5}, Supplementary Table 6).

We performed a GWAS for all 8 quantitative subphenotypes and 9 binary subphenotypes with at least 1,000 cases and calculated heritability and genetic correlation with BD and SCZ. Only two subphenotypes had significant h^2_{snp} estimates using LD-score regression(Bulik-Sullivan et al., 2015) both in BD: psychotic features in BD (h^2_{snp}=0.15, SE=0.06) and suicide attempt (h^2_{snp}=0.25, SE=0.1). Only psychotic features demonstrated a significant genetic correlation with SCZ (r_g=0.34, SE=0.13, p=0.009).

The significant genetic correlation demonstrates a genome-wide relationship between common variants contributing to SCZ risk and those contributing to psychotic features in BD cases. We tested whether the most significantly associated SCZ loci contributed directly to psychotic features in BD. One hundred of the 105 autosomal genome-wide significant SCZ SNPs previously published(Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014) were in our dataset after QC and 60 were in the same direction of effect for risk of psychotic features in BD (p=0.028, one-sided binomial-test).

Discussion
Here we present a genetic dissection of bipolar disorder and schizophrenia from over 100,000 genotyped subjects. Consistent with earlier results (Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013), we found extensive genetic sharing between these two disorders, identifying 114 genome-wide significant loci contributing to both disorders of which 32 are novel. These findings point to the relevance of neuronal and synaptic biology for the shared genetic substrate of these disorders. However, despite this degree of sharing, we identified several loci that significantly differentiated between the two disorders, having opposite directions of effect. We also found polygenic components that significantly correlated from one disorder to symptoms of the other.

Two GWS loci were identified from the case only SCZ versus BD analysis providing opportunities to inform the underlying biological distinctions between BD and SCZ. The most significant locus implicates DARS2 (coding for the mitochondrial Aspartate-tRNA ligase) which is highly expressed in the brain and significantly regulated by the most significant SNP rs56355601 ($p_{\text{eQTL}}=2.5 \times 10^{-11}$). Homozygous mutations in DARS2 are responsible for leukoencephalopathy with brainstem and spinal cord involvement and lactate elevation (LBSL), which was characterized by neurological symptoms such as psychomotor developmental delay, cerebellar ataxia and delayed mental development (Yamashita et al., 2013, p. 2). Based on methylation analysis from the prefrontal cortex of stress models (rats and monkeys) and from peripheral samples (in monkeys and human newborns), DARS2, among others, has been suggested as a potential molecular marker of early-life stress and vulnerability to psychiatric disorders (Luoni et al., 2016). The second most significant locus implicates CSEIL, a nuclear transport factor that plays a role in cellular proliferation as well as in apoptosis (Bera et al., 2001).
Intronic SNPs in CSE1L have been associated with subjective well-being (Okbay et al., 2016) and, nominally to antidepressant response (Li et al., 2016). More interestingly, CSE1L is a potential target gene of miR-137, one of the well-known schizophrenia risk loci (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014), which is able to negatively regulate CSE1L by interacting with complementary sequences in the 3’ UTR of CSE1L (Li et al., 2013). Although falling short of genome-wide significance, the third most significant locus implicates ARNTL (Aryl Hydrocarbon Receptor Nuclear Translocator Like), which is a core component of the circadian clock. ARNTL has been previously hypothesized for relevance in bipolar disorder, (Yang et al., 2008) although human genetic evidence is limited (Byrne et al., 2014).

The ability to generate transcriptional data on multiple tissues across many individuals using RNA-sequencing has provided detailed information on the role common variants play in regulating expression of specific genes in specific tissues. These eQTLs can be integrated with the genetic association data from GWAS to inform on the relationship between variant association and variant regulation of expression for each gene. Performing this integration, we identified a third genome-wide significant finding in DCAKD. The gene codes for Dephospho-CoA Kinase Domain Containing protein, a member of the human postsynaptic density proteome from human neocortex (Bayés et al., 2011). In the mouse cortical synaptoproteome DCAKD is among the proteins with the highest changes between juvenile postnatal days and adult stage, suggesting a putative role in brain development (Gonzalez-Lozano et al., 2016; Moczulska et al., 2014). Discerning between pleiotropy (variant independently regulates expression and alters risk to disease) from causality (variant regulates expression which thereby alters risk to disease)
through statistical analysis alone is difficult, this analytical approach is stringent in excluding loci where colocalised SNP-phenotype and SNP-expression associations may reflect confounding driven by linkage disequilibrium (LD) (one variant regulates expression and a different variant alters risk but the variants in the region are in LD). Hence, this approach utilizes currently available data to prioritize genes, including direction of effect, for functional follow-up. These analyses will become more powered with increased sample sizes for both phenotype and eQTL data sets.

Performing pathway analysis based on the full association results shows enrichment of genes involved in response to potassium ions, including potassium voltage-gated channel subfamily members and a number of genes regulated by cellular potassium concentration. This is in line with previous genetic evidence pointing to a key etiologic role of potassium channels, in particular, in BD (Judy and Zandi, 2013), which could be explained by their role in multiple neurobiological mechanisms involved in the development of psychiatric disorders such as regulation of the dopaminergic circuits, synaptic plasticity, and myelination (Balaraman et al., 2015).

We further assessed the contribution of regions of the genome to each disorder through joint regional association and heritability estimation. These results point to an additional locus that may contribute differentially to liability to BD and SCZ. The region on chr19 shows overlapping association peaks that are driven by independent causal variants for each disorder. Both variants significantly regulate the same gene GATAD2A but in opposite directions. GATAD2A is a transcriptional repressor, which is targeted by MBD2 and is involved in methylation-dependent
gene silencing. The protein is part of the large NuRD (nucleosome remodeling and deacetylase) complex, for which also HDAC1/2 are essential components. NurD complex proteins have been associated with autism (Li et al., 2015). Their members, including GATAD2A, display preferential expression in fetal brain development (Li et al., 2015) and in recent work has been implicated in SCZ through open chromatin (Fullard et al., n.d.). Further, p66α (mouse GATAD2A) was recently shown to participate in memory preservation through long-lasting histone modification in hippocampal memory-activated neurons (Ding et al., 2017). SNP-heritability appears to be consistently shared across regions and chromosomes between these two disorders. Regions with GWS loci often explain higher proportions of heritability as expected. When looking at the effect on heritability of the presence of a GWS locus in the other disorder, we identified a significant increase in BD heritability for regions containing a GWS locus for SCZ but no significant increase in SCZ heritability in regions having a BD one. This result suggests a directionality to the genetic sharing of these disorders with a larger proportion of BD loci being specific to BD. However, we cannot exclude that the asymmetry of results may reflect less power of discovery for BD than SCZ. The degree to which power and subphenotypes contribute to this result requires further examination.

We note that as with nearly all GWAS findings, the calculated population-based effect sizes of the variants identified here are small and independently explain only a modest fraction to the heritability of these disorders. The identification of these variants is dependent on the ability to have highly accurate allele frequency estimates that can only be ascertained from large sample sizes. As sample sizes get larger the power to identify variants of smaller effect increases meaning that increasing sample size results in the identification of variants of smaller effect.
However, a small population effect size does not exclude the possibility of a substantially larger
effect on molecular phenotypes nor does it preclude the utility of association regions in
understanding biology or having a clinical impact. Efforts following up GWAS results to date
have demonstrated the value of these findings in pointing to genes that can aid in understanding
the underlying biology of the trait (Claussnitzer et al., 2015; Mohanan et al., 2018; Sekar et al.,
2016). Further, there is a clear relationship between GWAS results of a phenotype and gene
targets of drugs that treat that phenotype pointing to the potential for improved therapeutic
understanding (Nelson et al., 2015; Ruderfer et al., 2016). A major challenge of GWAS is the
sheer number of findings and the substantial time/cost required for functional follow up of these
findings in the classical paradigms used for genes causal for monogenic disorders. In silico
bioinformatic analyses (such as SMR used here) that integrate GWAS results with ‘omics data
(transcription, protein, epigenetic, etc.) have the potential to put a clearer biological focus on
GWAS results. Such analyses can become more complex as more reference omics data sets (with
genome-wide genotyping) become available. Additional analytical efforts will be required to
facilitate the transition from GWAS to biology but substantial data has shown there is much to be
learned from these variants despite their small effects (Visscher et al., 2017).

We have now identified multiple genomic signatures that correlate between one disorder and a
clinical symptom in the other disorder, illustrating genetic components underlying particular
symptom dimensions within these disorders. Medical symptoms, including those seen in
psychiatric disorders, can manifest through a multitude of causes. The classic example often used
is headache for which many different paths lead to the same symptom. Psychiatric symptoms
also have many potential causes. For example, symptoms of psychosis can be the result of highly
heritable diseases such as BD and SCZ but also infectious and neurodegenerative diseases, sleep/sensory deprivation or psychedelic drugs. Demonstrating a shared biological underpinning to these symptoms suggests they could be treated through modulating the same pathway. As previously shown, we find a significant positive correlation between the PRS of BD and manic symptoms in SCZ. We also demonstrate that BD cases with psychotic features carry a significantly higher SCZ PRS than BD cases without psychotic features and this result is not driven by the schizoaffective BD subtype. Further, we show that increased PRS is associated with more severe illness. This is true for BD with psychotic features having increased SCZ PRS, earlier onset BD having higher SCZ PRS and cases with higher BD+SCZ PRS having a larger number of hospitalizations. We demonstrated that psychotic features within BD is a heritable trait and GWS loci for SCZ have a consistent direction of effect in psychotic features in BD, demonstrating the potential to study psychosis more directly to identify variants contributing to that symptom dimension.

This work illustrates the utility of genetic data, in aggregate, at dissecting symptom heterogeneity among related disorders and suggests that further work could aid in characterizing patients for more personalized treatment. Genetic risk scores have demonstrated their ability to inform and predict pathology(Cleynen et al., 2016) and more recently have been shown to be able to identify patients with risk equivalent to monogenic variants(Khera et al., 2017). In psychiatry, we lack objective biological measurements (biomarkers) with which to assess the ability of a genetic signature to predict or inform. Lacking diagnostic pathology for psychiatric disorders leaves a genuine opportunity for the genetics to drive diagnosis and treatment to a much larger degree than in other domains. One potential model assumes that each individual has
a quantitative loading of a series of symptom dimensions (i.e. manic, psychotic, cognitive, etc.)
and that these symptoms can be assessed at the genetic level to characterize a patient’s
dysfunction and used to inform disease course and optimal treatment. Making this a reality will
require more detailed information on disease course and outcomes. For example, if treatment
response data existed for these samples one could ask whether a genetic loading for psychosis
was correlated with response to treatment. Initial work has already shown the potential of this
approach using a SCZ PRS to inform lithium response in BD (Amare et al., 2018). Ultimately,
the goal will be to quantify multiple genetic loadings of each individual’s illness and use those
measures to inform treatment based on the outcomes of previous individuals with similar
profiles.

In conclusion, we present a detailed genetic dissection of BD and SCZ pointing to substantial
shared genetic risk but also demonstrating that specific loci contribute to the phenotypic
differences of these disorders. We show that genetic risk scores can correspond to symptoms
within and across disorders. Finally, we present data that points to these disorders being neither
independent nor the same but sharing particular symptom dimensions that can be captured from
the genetics and used to characterize patients to ultimately inform diagnosis and treatment.

**Author Contributions:**

DMR, PS and KSK managed and organized the group. DMR, SR, JB, EAS, JMWP, NM, AWC,
APSO, LMOL and VT contributed to analyses. Subphenotype collection and organization was
led by AM and AHF. Initial manuscript was drafted by DMR, ED, ADF, SP, JLK. Manuscript
contributions and interpretation of results was provided by DMR, ED, SHL, MCO, PFS, RAO,
NRW, PS and KSK. The remaining authors contributed to the recruitment, genotyping, or data
processing for the contributing components of the study. All other authors saw, had the
opportunity to comment on, and approved the final draft.

**Consortia**

**Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium**

Douglas M Ruderfer, Stephan Ripke, Andrew McQuillin, James Boocock, Eli A Stahl, Jennifer M Whitehead,
Pavlides, Niamh Mullins, Alexander W Charney, Anil P S Ori, Loes M Olde Loohuis, Enrico Domenici, Arianna Di
Florio, Sergi Papiol, Janos L. Kalman, Vassily Trubetskoy, Rolf Adolfsson, Ingrid Agartz, Esben Agerbo,
Huda Akil, Diego Albani, Margot Albus, Martin Alda, Madeline Alexander, Ney Alliey-Rodriguez, Thomas D Als,
Farooq Amin, Adebayo Anjorin, Maria J Arranz, Swapnil Awashti, Silviu A Bacanu, Judith A Badner, Marie
Baekvad-Hansen, Steven Bakker, Gavin Band, Jack D Barchas, Ines Barroso, Nicholas Bass, Michael Bauer,
Bernhard T Baune, Martin Begemann, Celine Bellenguez, Richard A Belliveau Jr, Frank Bellivier, Stephan Bender,
Judit Bene, Sarah E Bergen, Wade H Berrettini, Elizabeth Bevilacqua, Joanna M Biernacka, Tim B Bigdeli, Donald
W Black, Hannah Blackburn, Jenefer M Blackwell, Douglas HR Blackwood, Carsten Bocker Pedersen, Michael
Boehnke, Marco Boks, Anders D Borglum, Elvira Bramon, Gerome Breen, Matthew A Brown, Richard Bruggeman,
Nancy G Buccola, Randy L Buckner, Monika Budde, Brendan Bulik-Sullivan, Suzannah J Bumpstead, William
Bunney, Margit Burmeister, Joseph D Buxbaum, Jonas Bybjerg-Grauholm, William Byerley, Wiepke Cahn,
Guiqing Cai, Murray J Cairns, Dominique Campion, Rita M Cantor, Vaughan J Carr, Noa Carrera, Juan P Casas,
Miguel Casas, Victoria V Catts, Pablo Evelo, Kimberly D Chambert, William Cohen, Nadine Cohen, Jonathan I Coleman, David
A Collier, Paul Cormican, William Coryell, Nicholas Craddock, David W Craig, Benedikt Crespo-Facorro, James J Crowley, Cristiana Cruz-est, David Curtis, Piotr M
Czerski, Anders M Dale, Mark J Daly, Udo Dannlowski, Ariel Davvasi, Michael Davidson, Kenneth L Davis,
Christianna A de Leeuw, Franziska Degenhardt, Jurgen Del Favero, Lynn E DeLisi, Panos Deloukas, Ditte Demontis,
J Raymond DePaulo, Marta di Forti, Dimitris Dikeos, Timothy Dinan, Srdjan Djurovic, Amanda L Dobbny, Peter
Donnelly, Gary Donohoe, Elodie Drapeau, Serge Dronov, Juba Duan, Frank Durdbridge, Audrey Duncanson,
Howard Edenberg, Sarah Edkins, Hannelore Ehrenreich, Peter Eichhammer, Torbjorn Elvsashagen, Johan Eriksson,
Valentina Escott-Price, Tonu Esko, Laurent Essioux, Bruno Etain, Chun Chieh Fan, Kai-How Farh, Martillas S
Farrell, Matthew Flickinger, Tatiana M Foroud, Liz Forty, Josef Frank, Lude Franke, Christine Fraser, Robert
Freedman, Colin Freeman, Nelson B Freimer, Joseph I Friedman, Menachem Fromer, Mark A Frye, Janice M
Fullerton, Katrin Gade, Julie Garnham, Helena A Gaspar, Pablo V Gejman, Giulio Genovese, Lyudmila Georgiev,
Claudia Giambartolomei, Eleni Giannoulatou, Ina Giegling, Michael Gill, Matthew Gillman, Marianne Giortz
Pedersen, Paola Giusti-Rodriguez, Stephanie Godard, Fernando Goes, Jacqueline J Goldstein, Srihari Gopal, Scott D
Gordon, Katherine Gordon-Smith, Jacob Gratten, Emma Gray, Elaine K Green, Melissa J Green, Tiffany A
Greenwood, Maria Grigorou-Serbanescu, Jakob Grove, Weihua Guan, Hugh Gurling, Jose Guzman Parra, Rhiannon
Gwilliam, Lieuwe de Haan, Jeremy Hall, Mei-Hua Hall, Christian Hammer, Naomi Hammond, Marline Lhamshire,
Mark Hansen, Thomas Hansen, Vahram Haroutunian, Annette M Hartmann, Joanna Hauser, Martin Hautzinger, Urs
Heilbronner, Garrett Hellenthal, Frans A Henskens, Stefan Herm, Maria Hipolito, Noel J Hirschhorn, Per
Hoffmann, Mads V Hollegaard, David M Hougaard, Hailliang Huang, Laura Huckins, Christina M Hultman, Sarah E
Hunt, Masashi Ikeda, Nakao Iwata, Conrad Iyegbe, Tao Li, Kung-Yee Liang, Jennifer Liddle, Jeffrey Lieberman, Svetlana Limborska, Kwang Lin, Don H
Linszen, Jolanta Lissowska, Chunyu Liu, Jianjun Liu, Jouko Lonnqvist, Carmel M Lowndie, Jan Lubinski,
Acknowledgements

The work of the contributing groups was supported by numerous grants from governmental and charitable bodies as well as philanthropic donation. Specifically, NRW (NHMRC 1078901, 1087889).

Declaration of Interests
The authors declare no competing interests.

References


https://doi.org/10.1038/ng.3211


splicing and memory preservation. Nat. Neurosci. advance online publication.

https://doi.org/10.1038/nn.4536

Fromer, M., Pocklington, A.J., Kavanagh, D.H., Williams, H.J., Dwyer, S., Gormley, P.,
Georgieva, L., Rees, E., Palta, P., Ruderfer, D.M., Carrera, N., Humphreys, I., Johnson,
J.S., Roussos, P., Barker, D.D., Banks, E., Milanova, V., Grant, S.G., Hannon, E., Rose,
S.A., Chambert, K., Mahajani, M., Scolnick, E.M., Moran, J.L., Kirov, G., Palotie, A.,
McCarroll, S.A., Holmans, P., Sklar, P., Owen, M.J., Purcell, S.M., O’Donovan, M.C.,
2014. De novo mutations in schizophrenia implicate synaptic networks. Nature 506,
179–184. https://doi.org/10.1038/nature12929

Fromer, M., Roussos, P., Sieberts, S.K., Johnson, J.S., Kavanagh, D.H., Perumal, T.M., Ruderfer,
D.M., Oh, E.C., Topol, A., Shah, H.R., Klei, L.L., Kramer, R., Pinto, D., Gümüş, Z.H., Cicek,
A.E., Dang, K.K., Browne, A., Lu, C., Xie, L., Readhead, B., Stahl, E.A., Xiao, J., Parvizi, M.,
Hamamsy, T., Fullard, J.F., Wang, Y.-C., Mahajani, M.C., Derry, J.M.J., Dudley, J.T.,
Zhang, B., Sullivan, P.F., Chess, A., Purcell, S.M., Shinobu, L.A., Mangravite, L.M.,
Toyoshiba, H., Gur, R.E., Hahn, C.-G., Lewis, D.A., Haroutunian, V., Peters, M.A., Lipska,
B.K., Buxbaum, J.D., Schadt, E.E., Hirai, K., Roeder, K., Brennand, K.J., Katsanis, N.,
https://doi.org/10.1038/nn.4399

Fullard, J.F., Giambartolomei, C., Hauberg, M.E., Xu, K., Voloudakis, G., Shao, Z., Bare, C,
chromatin profiling of human postmortem brain infers functional roles for non-

gene-based association method for mapping traits using reference transcriptome
data. Nat. Genet. 47, 1091–1098. https://doi.org/10.1038/ng.3367

Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C,
Plagnol, V., 2014. Bayesian Test for Colocalisation between Pairs of Genetic
https://doi.org/10.1371/journal.pgen.1004383

Gonzalez-Lozano, M.A., Klemmer, P., Gebuis, T., Hassan, C., van Nierop, P., van Kesteren, R.E.,
Smit, A.B., Li, K.W., 2016. Dynamics of the mouse brain cortical synaptic proteome
https://doi.org/10.1038/srep35456

Green, E.K., Rees, E., Walters, J.T.R., Smith, K.-G., Forty, L., Grozeva, D., Moran, J.L., Sklar, P.,
Ripke, S., Chambert, K.D., Genovese, G., McCarroll, S.A., Jones, I., Jones, L., Owen, M.J.,
O’Donovan, M.C., Craddock, N., Kirov, G., 2016. Copy number variation in bipolar

Neuropsychopharmacology 40, 252–254. https://doi.org/10.1038/npp.2014.216

Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B.W.J.H., Jansen, R., de Geus, E.J.C.,
Boomsma, D.I., Wright, F.A., Sullivan, P.F., Nikkola, E., Alvarez, M., Civelek, M., Lusis,
A.J., Lehtimäki, T., Raitoharju, E., Kähönen, M., Seppälä, I., Raitakari, O.T., Kuusisto, J.,
large-scale transcriptome-wide association studies. Nat. Genet. 48, 245–252. https://doi.org/10.1038/ng.3506


from GWAS and eQTL studies predicts complex trait gene targets. Nat. Genet. 48, 481–487. https://doi.org/10.1038/ng.3538

Figure Legends

Figure 1. Associated Genomics Loci Shared and Divergent Between BD and SCZ
a) Odds ratios (OR) from independent data sets of BD (blue) and SCZ (red) for each of the 114 genome-wide significant variants in the BD and SCZ vs controls GWAS. b) Manhattan plot for SCZ vs BD GWAS.

Figure 2. Genomic Region with Disorder Independent Causal Variants in BD and SCZ
Regional association plot of chr19 18.4 - 20 Mb. Independent GWAS of SCZ (top) and BD (bottom) are displayed along with genes in the region. Each point is a SNP with the most significant SNP labeled and colored purple, other points are colored by LD with most significant SNP. Despite overlapping regional association these loci are independent across disorders and conditional analysis in both directions yields equivalent significance.

Figure 3. Polygenic Risk Score Dissection of Clinical Symptom Dimensions
Effect size (calculated by dividing regression estimate by standard error) from regression analysis including ancestry covariates for each subphenotype and PRS for BD (x-axis) and SCZ
(y-axis). Point size represents \(-\log_{10}(p\text{-value})\) with SCZ (red) and BD (blue). Numbered subphenotypes are 1) comorbid migraine, 2) panic attacks 3) suicide attempt 4) mixed states 5) rapid cycling 6) comorbid eating disorder 7) comorbid OCD 8) year of birth 9) suicide ideation 10) panic disorder 11) number of suicide attempts 12) depressive symptoms (SCZ) 13) episodes depressive 14) episodes total 15) positive symptoms (SCZ) 16) irritable mania 17) age of onset depression 18) family history 19) episodes mixed mania 20) unipolar mania 21) alcohol substance dependence 22) age of onset mania 23) age at interview 24) number of hospitalizations. All subphenotypes are in BD except those labeled (SCZ).

Table Legends

Table 1. Most Significant Associated Loci from SCZ vs BD GWAS
Association results for the five most significant variants in the SCZ vs BD GWAS with the top two being genome-wide significant. Each variant includes results from the independent BD vs controls and SCZ vs controls GWAS and the comparable p-value from a heterogeneity test when performing a two cohort meta-analysis of SCZ and BD.

Table 2. Complete Results of Polygenic Risk Score Dissection Analysis
Polygenic scoring results of all four GWAS phenotypes (BD+SCZ vs controls, BD vs controls, SCZ vs controls and SCZ vs BD) and 24 subphenotypes from BD and 4 subphenotypes from SCZ, rows without case/control counts are quantitative measures. Significance and effects are from regression analysis of subphenotype on PRS including principal components of ancestry and site as covariates. Effect is the regression estimate divided by the standard error.
Supplementary Figure Legends

Figure S1. Related to Figure 1b. Regional Association Plot and Forest Plot for the First Genome-wide Significant Hit in the SCZ vs BD GWAS.

Figure S2. Related to Figure 1b. Regional Association Plot and Forest Plot for the Second Genome-wide Significant Hit in the SCZ vs BD GWAS.

Figure S3. Related to Summary-data-based Mendelian Randomization. Detailed Association of DCAKD from SMR.

Results at the DCAKD locus from SMR analysis of SCZ vs BD. Top plot, brown dots represent the $P$ values for SNPs from SCZ vs BD GWAS, diamonds represent the $P$ values for probes from the SMR test. Bottom plot, the eQTL $P$ values of SNPs from the Westra study for the ILMN_1811648 probe tagging DCAKD. The top and bottom plots include all the SNPs available in the region in the GWAS and eQTL summary data, respectively, rather than only the SNPs common to both data sets. Highlighted in red is the gene (DCAKD) that passed the SMR and HEIDI tests.

Figure S4. Related to Regional SNP-heritability estimation. Heritability Estimates for BD and SCZ in Genome-wide Significant Regions of BD and SCZ.

Regional SNP-heritability estimates for SCZ and BD stratified by whether the region contains the most significant variant in a genome-wide significant locus in BD, SCZ, neither or both.
CONTACT FOR REAGENT AND RESOURCE SHARING

Summary statistics from GWAS are publically available at https://www.med.unc.edu/pgc/results-and-downloads/downloads. Genotype and phenotype data use is restricted and governed by the Psychiatric Genetics Consortium. Further information and requests for analytical results or additional information should be directed to and will be fulfilled by the Lead Contact, Douglas Ruderfer (douglas.ruderfer@vanderbilt.edu).

SUBJECT DETAILS

Genotyped Sample Description

SCZ samples are a substantial subset of those analyzed previously (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). BD samples are the newest collection from Psychiatric Genomics Consortium Bipolar Disorder Working Group (Stahl et al., 2017). To ensure independence of the data sets, individuals were excluded until no individual showed a relatedness (pihat) value greater than 0.2 to any other individual in the collection, while preferentially keeping the case over the control for case-control related pairs. In total 1,795 BD cases, 1,165 SCZ cases and 27,274 controls were removed (most of which were previously known), leaving 20,129 BD cases 33,426 SCZ cases and 54,065 controls for the final meta-analysis.

For analyses directly comparing BD and SCZ, we matched cases from both phenotypes on genotyping platform and ancestry, resulting in 15,270 BD cases versus 23,585 SCZ cases.
Hence, we were able to match 76% of BD cases and 71% of SCZ cases for this case vs case analysis. All samples were collected with approval from respective IRBs and informed consent was obtained.

Sub-phenotype Description

BD sub-phenotypes were collected by each study site using a combination of diagnostic instruments, case records and participant interviews. Ascertainment details for each study site are described in the supplementary data of the PGC Bipolar Working Group paper (Stahl et al., 2017). The selection of phenotypes for collection by this group was determined by literature searches in order to determine phenotypes with prior evidence for heritability. It was further refined dependent on the availability of phenotype data across a range of study sites and the consistency by which the phenotypes were defined. Schizophrenia subphenotypes are the same as described previously (Ruderfer et al., 2014) but in a larger proportion of patients.

METHOD DETAILS

QUANTIFICATION AND STATISTICAL ANALYSIS

Quality Control, Imputation, Association Analysis and Polygenic Risk Score Testing

Quality control and imputation were performed on each of the study cohort datasets (n=81), according to standards established by the Psychiatric Genomics Consortium (PGC). The quality control parameters for retaining SNPs and subjects were: SNP missingness < 0.05 (before sample removal); subject missingness (p < 0.02); autosomal heterozygosity deviation (| Fhet | <
SNP missingness < 0.02 (after sample removal); difference in SNP missingness between cases and controls < 0.02; and SNP Hardy-Weinberg equilibrium ($p > 10^{-6}$ in controls or $p > 10^{-10}$ in cases). Genotype imputation was performed using the pre-phasing/imputation stepwise approach implemented in IMPUTE2 (Howie et al., 2011) / SHAPEIT (Delaneau et al., 2013) (chunk size of 3 Mb and default parameters). The imputation reference set consisted of 2,186 phased haplotypes from the full 1000 Genomes Project dataset (August 2012, 30,069,288 variants, release “v3.macGT1”), all variants align to human genome build 19 (hg19). After imputation, we used the best guess genotypes (genotype probability > 0.8), for further robust relatedness testing and population structure analysis. Here we required very high imputation quality (INFO > 0.8) and low missingness (<1%) for further quality control. After linkage disequilibrium (LD) pruning ($r^2 < 0.02$) and frequency filtering (MAF > 0.05), there were 14,473 autosomal SNPs in the data set. Principal component estimation was done with the same collection of autosomal SNPs. We tested the first 20 principal components for phenotype association (using logistic regression with study indicator variables included as covariates) and evaluated their impact on the genome-wide test statistics using $\lambda$. Thirteen principal components namely 1,2,3,4,5,6,7,8,10,12,15,18,20 were included in all association analyses ($\lambda=1.45$). Analytical steps were repeated for SCZ vs BD analysis. We performed four main association analyses (Figure 1), i.e. (i) GWAS of BD and SCZ as a single combined case phenotype, as well as disorder-specific GWAS using independent control sets in (ii) BD cases vs BD controls and (iii) SCZ cases vs SCZ controls, and (iv) association analysis of SCZ cases vs BD cases. For all GWS loci from the GWAS of BD and SCZ vs controls we identified any GWS loci within 1Mb from the extent of the locus in the previously published PGC SCZ vs controls (Schizophrenia Working Group of the Psychiatric Genomics).
Consortium, 2014) and the most recent PGC GWAS of BD vs controls (Stahl et al., 2017) and performed conditional analysis. Specifically, we transformed the genotype probabilities of the disease variant into dosages and used it as an additional covariate for the association analysis for the BD+SCZ vs controls index SNP. This was done within each cohort and an OR based inverse SE weighted meta-analysis was performed for the final result. All datasets were included except for those with trios.

**Summary-data-based Mendelian Randomization (SMR)** (Zhu et al., 2016)

SMR is a method that integrates summary level GWAS data with gene expression quantitative trait loci (eQTL) identified in independent data sets. This integration aims to identify variants that have pleotropic effects on expression of a given gene and the phenotype. While significant findings may indeed reflect a causal path from variant to phenotype through expression, it is impossible to discern statistically between pleiotropy and causality. However, the method can remove linkage as driving the result, and uses the data available to prioritise amongst genes in genomic regions that show association with disease. We used SMR as a statistical fine-mapping tool applied to the SCZ vs BD GWAS results to identify loci with strong evidence of causality via gene expression. SMR analysis is limited to significant (FDR < 0.05) cis SNP-expression quantitative trait loci (eQTLs) with MAF > 0.01. eQTLs passing these thresholds were combined with GWAS results in the SMR test, with significance (p_{SMR}) reported at a Bonferroni-corrected threshold for each eQTL data set. The eQTL architecture may differ between genes. For example, through LD, many SNPs can generate significant associations with the same gene, but in some instances multiple SNPs may be independently associated with the expression of a gene. After identification of significant SNP-expression-trait association through the SMR test, a
follow-up heterogeneity test aims to prioritize variants by excluding regions for which there is conservative evidence for multiple causal loci ($p_{\text{HET}} < 0.05$). SMR analyses were conducted using eQTL data from whole peripheral blood (Westra et al., 2013), dorsolateral prefrontal cortex generated by the CommonMind Consortium, and 11 brain sub-regions from the GTEx consortium (Consortium, 2015).

**Regional joint GWAS**

Summary statistic Z-scores were calculated for each marker in each of the four main GWAS results, using the logistic regression coefficient and its standard error. Rare SNPs (MAF < 0.01), and SNPs with a low INFO score (< 0.3) in either dataset were removed. The causal variant relationships between SCZ and BD were investigated using the Bayesian method software pw-gwas (v0.2.1), with quasi-independent regions determined by estimate LD blocks in an analysis of European individuals (n=1,703) (Berisa and Pickrell, 2015; Pickrell et al., 2016). Briefly, pw-gwas takes a Bayesian approach to determine the probability of five independent models of association. (1) There is no causal variant in BD or SCZ; (2) a causal variant in BD, but not SCZ; (3) a causal variant in SCZ, but not BD; (4) a shared causal variant influencing both BD and SCZ; (5) two causal variants where one influences BD, and one influences SCZ (Figure 2). The posterior probability of each model is calculated using model priors, estimated empirically within pw-gwas. Regions were considered to support a particular model when the posterior probability of the model was greater than 0.5.

**Regional SNP-heritability estimation**
We calculated local SNP-heritability independently for SCZ and BD using the Heritability Estimator from Summary Statistics (HESS) software (Shi et al., 2016) for each of the independent regions defined above. The sum of these regional estimates is the total SNP-heritability of the trait. To calculate local SNP-heritability HESS requires reference LD matrices representative of the population from which the GWAS samples were drawn. We utilized the 1000 genomes European individuals as the reference panel (The 1000 Genomes Project Consortium, 2015). Unlike pw-gwas (Pickrell et al., 2016), HESS does not assume that only one causal variant can be present in each region.

DATA AND SOFTWARE AVAILABILITY

Summary statistics from GWAS are publically available at https://www.med.unc.edu/pgc/results-and-downloads/downloads.