**Genomic dissection of bipolar disorder and schizophrenia in 50k cases, 50k controls and 28 subphenotypes**

*Preliminary author list and unfinished affiliations, needs to include BD and SCZ working groups*

Douglas M. Ruderfer, Stephan Ripke, Andrew McQuillin, James Boocock, Eli Stahl, Jenny Pavlides, Niamh Mullins, Alexander Charney, Enrico Domenici, Arianna Di Florio, Sergi Papiol, Janos L. Kalman, Aiden P. Corvin, Ayman Fanous, Marcella Rietschel, Elliot Gershon, Thomas G. Schulze, Alfredo B Cuellar-Barboza, Andreas Forstner, Peter Holmans, John Nurnberger, Ole Andreassen, Hong Lee, Michael O'Donovan, Patrick Sullivan, Roel Ophoff, Naomi Wray, Pamela Sklar*, Kenneth S. Kendler*

**Affiliations**
1. Division of Genetic Medicine, Departments of Medicine; Department of Psychiatry; Department of Biomedical Informatics; Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN.
2. Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts; Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts.
3. Molecular Psychiatry Laboratory; Research Department of Mental Health Sciences; University College London Medical School; Windeyer Institute of Medical Sciences; London; England; UK
4. Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, USA.
5. Division of Psychiatric Genomics, Icahn School of Medicine at Mount Sinai; Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai
6. Queensland Brain Institute, The University of Queensland, Brisbane, Australia
7. Laboratory of Neurogenomic Biomarkers, Centre for Integrative Biology (CIBIO), University of Trento, Trento
8. MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, UK;
9. Institute of Psychiatric Phenomics and Genomics (IPPG), University Hospital, Ludwig Maximilian University, Munich, Germany
10. Department of Psychiatry and Psychotherapy, Ludwig Maximillian University, Munich, Germany
11. Neuropsychiatric Genetics Research Group, Department of Psychiatry, Trinity College Dublin, Dublin 8, Ireland
12. Department of Psychiatry, Institute for Genomic Health SUNY Downstate Medical Center Brooklyn, NY
13. Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health Mannheim, University Heidelberg, J5, 68159 Mannheim, Germany
14. Department of Psychiatry, University of Chicago, Chicago, Illinois
15. Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Mannheim, Germany
16. Department of Psychiatry and Behavioral Sciences, The Johns Hopkins University, Baltimore, MD
17. Human Genetics Branch, NIMH, NIH, Bethesda, MD
18. Universidad Autonoma de Nuevo Leon, Department of Psychiatry, Monterrey, México. Department of Psychiatry, Mayo Clinic, MN
19. Institute of Human Genetics, University of Bonn, D-53127 Bonn, Germany
20. Department of Genomics, Life and Brain Center, D-53127 Bonn, Germany; Applied Molecular Genomics Unit, VIB Department of Molecular Genetics,
21. Department of Psychiatry, Indiana University School of Medicine; Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, Indiana

22. Institute of Clinical Medicine, ENBREC Group, University of Oslo, N-0407 Oslo, Norway; Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway
23. University of North Carolina at Chapel Hill, Karolinska Institutet
24. Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, California
25. University Medical Center Utrecht, Department of Psychiatry, Rudolf Magnus Institute of Neuroscience, 3584 Utrecht, The Netherlands
26. Department of Genetics and Genomics, Icahn School of Medicine at Mount Sinai; Friedman Brain Institute, Icahn School of Medicine at Mount Sinai. New York, NY
27. Virginia Institute for Psychiatric and Behavioral Genetics and Department of Psychiatry, Virginia Commonwealth University, Richmond, VA, USA.

**Corresponding author:** Douglas M. Ruderfer (douglas.ruderfer@vanderbilt.edu)

**Abstract**

Schizophrenia (SCZ) and bipolar disorder (BD) are highly heritable disorders that share a significant proportion of common risk variation. Understanding the genetic factors underlying the specific symptoms of these disorders will be crucial for improving diagnosis, intervention and treatment. In case-control data consisting of 53,555 cases (20,129 BD, 33,426 SCZ) and 54,065 controls, we identified 114 genome-wide significant loci (GWS) when comparing all cases to controls, of which 41 represented novel findings. Two genome-wide significant loci were identified when comparing SCZ to BD and a third was found when directly incorporating functional information. Regional joint association identified a genomic region of overlapping association in BD and SCZ with disease independent causal variants indicating a fourth region contributing differences between these disorders. Regional SNP-heritability analyses demonstrated that the estimated heritability of BD based on the SCZ GWS regions was significantly higher than that based on the average genomic region (91 regions, $p = 1.2 \times 10^{-6}$) while, the inverse was not significant (19 regions, p=0.89). Using our BD and SCZ GWAS we calculated polygenic risk scores and identified several significant correlations with: 1) SCZ

subphenotypes: negative symptoms (SCZ, $p=3.6\times10^{-6}$) and manic symptoms (BD, $p=2\times10^{-5}$), 2) BD subphenotypes: psychotic features (SCZ $p=1.2\times10^{-10}$, BD $p=5.3\times10^{-5}$) and age of onset (SCZ $p=7.9\times10^{-4}$). Finally, we show that psychotic features in BD has significant SNP-heritability ($h^2_{snp}=0.15$, SE=0.06), significant genetic correlation with SCZ ($r_g=0.34$) and significant sign test result between SCZ GWAS and a GWAS of BD cases contrasting those with and without psychotic features ($p=0.0038$, one-side binomial test). For the first time, we have identified specific loci pointing to a potential role of 4 genes (*DARS2*, *ARFGEF2*, *DCAKD* and *GATAD2A*) that distinguish between BD and SCZ, providing an opportunity to understand the biology contributing to differing symptoms of these disorders. Our results provide the best evidence so far of genomic components divergent between BD and SCZ contributing directly to specific symptom dimensions.

**Introduction**

Bipolar disorder (BD) and schizophrenia (SCZ) are severe psychiatric disorders and among the leading causes of disability worldwide[1]. Both disorders have significant genetic components with heritability estimates ranging from 60-80%[2]. A genetic-epidemiological genetic study demonstrated a substantial overlap between these two disorders with a genetic correlation from common variation near 0.6-0.7 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)[3]. Despite shared genetics and symptomology, the current diagnostic systems[4,5], represent BD and SCZ as two distinct 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. Further,

premorbid cognitive impairment and reduced intelligence are more frequently and substantially affected in SCZ than BD[6]. The genetic contributors to these phenotypic distinctions have yet to be elucidated and could aid in understanding the underlying biology of their unique symptom dimensions.

While the shared genetic component is large, studies to date have identified key genetic architecture differences between these two disorders. A polygenic risk score created from a case only SCZ vs BD genome-wide association study (GWAS) significantly correlated with SCZ vs BD diagnosis in an independent sample[7], providing evidence that differences between the disorders also has a genetic basis. An enrichment of rare, moderate to highly penetrant copy number variants (CNVs) and *de novo* CNVs are seen in SCZ patients[8–12], while, the involvement of CNVs in BD is much less clear[13]. Although the role of *de novo* single nucleotide variants in BD and SCZ has been investigated in only a handful of studies so far, enrichment in pathways associated with the postsynaptic density has been reported for SCZ, but not BD[14,15]. Identifying disorder specific variants or quantifying the contribution of variation to specific symptom dimensions remains an open question. For example, previous work by this group has demonstrated that SCZ patients with greater manic symptoms had higher polygenic risk for BD[7]. Here, we utilize the largest collection of genotyped samples of BD and SCZ along with 28 subphenotypes to assess variants and genomic regions that contribute differentially to the disorders and to specific symptoms dimensions or subphenotypes within them.


**Methods**

**Sample Description**

SCZ samples are those analyzed previously[16]. BD samples are the newest collection from Psychiatric Genomics Consortium Bipolar Disorder Working Group (PGC BD2 paper). 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 2,181 BD cases, 1,604 SCZ cases and 27,308 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. In other words, we were able to match 76% of BD cases and 71% of SCZ cases.

**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 (PGC BD2 paper). 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 described previously but in a larger proportion of patients[7]

**Quality Control, Imputation, Association Analysis and Polygenic Risk Scoring**

Quality control and imputation were performed on each of the study cohorts (n=81) datasets 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 (| $F_{het}$ | < 0.2); 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[17] / SHAPEIT[18] (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"). After imputation, we used the best guess genotypes, 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 pruning ($r^2$ < 0.02) and frequency filtering (MAF > 0.05), there were 14,473 autosomal SNPs in the data set. Relatedness testing was done with PLINK[19] and pairs of subjects with pihat > 0.2 were identified and one member of each pair removed at random after preferentially retaining cases over controls. 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 λ. Thirteen principal components namely 1,2,3,4,5,6,7,8,10,12,15,18,20 were included in all association analyses (λ=1.45). Analytical steps were repeated for SCZ vs BD analysis.

We performed four main association analyses, i.e. (i) GWAS of BD and SCZ as a single combined case phenotype, disorder specific GWAS using independent control sets in (ii) BD cases vs BD controls and (iii) SCZ cases vs SCZ controls, as well as (iv) association analysis of SCZ cases vs BD cases.

**Summary-data-based Mendelian Randomization (SMR)[20]**

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 (eQTLs) and SNPs 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. Through linkage disequilibrium (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 ($p_{HET}$ < 0.05). SMR analyses were conducted using eQTL data from whole peripheral blood[21], dorsolateral prefrontal cortex generated by the CommonMind Consortium[8], and 11 brain sub-regions from the GTEx consortium[22].

**Regional joint GWAS**

Summary statistic Z-scores were calculated from the association p-value 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=1,702)[23,24]. 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. 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[25] 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 population from which the GWAS samples were drawn. We utilized the 1000 genomes European individuals as the reference panel[26]. Unlike pw-gwas[23], HESS does not assume that only one causal variant can be present in each region.

**Results**

**GWAS**

We performed association analysis of BD and SCZ as a combined phenotype, totaling 53,555 cases (20,129 BD, 33,426 SCZ) and 54,065 controls on 15.5 million dosages imputed from 1000 genomes phase 3[26]. Logistic regression was performed controlling for 13 components of ancestry, study sites and genotyping platform. One hundred and fourteen regions contained at least one variant of which the p-value was lower than our genome-wide significance (GWS) threshold of $p < 5 \times 10^{-8}$. Among these 114 loci, 41 had non-overlapping LD regions ($r^2 > 0.6$) with the largest and most recently performed single disease GWAS of SCZ[16] and BD (PGC BD2 paper). 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 that subsets of variants have larger effects in SCZ vs BD or vice versa.

To identify loci with distinguishing effects on BD and SCZ, we performed an association analysis on 23,585 SCZ cases and 15,270 BD cases matched for shared ancestry and genotyping platform (see Methods, Figure 1b Supplementary Figures 1-5, Supplementary Table 1). Two genome-wide significant loci were identified, the most significant was rs56355601 located on chromosome 1 at position 173,811,455 within an intron of *DARS2*. The second most significant locus was a four base indel 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. We sought to identify additional disease specific loci by incorporating expression information with association results to perform fine-

mapping and identify novel variants[27–30]. Here, we applied the summary-data-based Mendelian randomization (SMR) method[20] (see Methods) utilizing the cis-QTLs derived from peripheral blood[21], human dorsolateral prefrontal cortex (DLPFC)[31] from the Common Mind Consortium and 11 brain regions from the GTEx consortium[22]. 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 the SCZ vs BD ($p_{GWAS} = 2.8 \times 10^{-6}$) 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 DLFC, and $p_{eQTL} = 4.6 \times 10^{-15}$, $p_{SMR} = 6.0 \times 10^{-5}$ in GTEx cerebellar hemisphere) (Supplementary Table 2, Supplementary Figure 6) and shows no evidence of heterogeneity ($p_{HET} = 0.66$) which implies only a single causal variant in the region.

**Regional joint association**

We expanded our efforts to identify disorder specific genomic regions by jointly analyzing independent GWAS results from BD and SCZ[23]. Among 1,702 regions genome-wide (see Methods), 223 had a posterior probability of greater than 0.5 of having a causal variant in 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 zero were BD specific. Of note, the data estimated prior probability of having a BD specific region was 0.1% compared to 15% for SCZ, potentially a result of increased power from the larger SCZ samples size.

The 114 GWS SNPs from the combined BD and SCZ GWAS localized into 99 independent regions, of these 78 (79%) were shared with posterior probability of 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 (Supplementary Figure 7). The most significant variant in BD was rs111444407 (chr19:19358207, p = $8.67 \times 10^{-10}$) and for SCZ was rs2315283 (chr19:19480575, p=$4.41 \times 10^{-7}$). After conditioning on the most significant variant in the other disorder, the association signal of the most significant variant in BD and SCZ were largely unchanged (BD rs111444407 =$1.3 \times 10^{-9}$, SCZ rs2315283 p=$6.7 \times 10^{-5}$). We further calculated the probability of each variant in the region being causal for both BD and SCZ[32] 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[31] but in opposite directions (rs111444407 $p_{eQTL}$ = $6 \times 10^{-15}$, beta = 0.105; rs2315283 $p_{eQTL}$ = $1.5 \times 10^{-28}$, beta = -0.11).

**Regional SNP-heritability estimation**

Across the genome, regional SNP-heritabilities ($h^2_{snp}$) were estimated separately for SCZ and BD[25] and were found to be moderately correlated (r=0.25). We next defined risk regions as those containing the most associated SNP for a GWS locus. In total, there were 101 SCZ risk regions from the 105 autosomal GWS loci reported previously[16] and 29 BD risk regions from 30 GWS

loci reported in a companion paper (PGC BD2 paper). 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 $h^2_{snp}$ by whether a region was a risk region in one disorder, none or both (Figure 2). 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 $h^2_{snp}$ in SCZ, as expected (p = $1.1\text{x}10^{-22}$), but also in BD (p = $1.2\text{x}10^{-6}$). In risk regions specific to BD (n=19), significantly increased regional $h^2_{snp}$ 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 $h^2_{snp}$ in both disorders, as expected (BD p = $5.3\text{x}10^{-5}$, SCZ p = 0.006), compared to non-risk regions. However, we observed a significant increase in BD $h^2_{snp}$ in shared risk regions compared to BD risk regions (BD p = 0.003) but not SCZ $h^2_{snp}$ for shared risk regions compared to SCZ risk regions (p = 0.62). Using a less stringent p-value threshold for defining risk regions (p < $5\text{x}10^{-6}$), thereby substantially increasing the number of regions, resulted in similar results (Supplementary Figure 8). Seven regions contributed to substantially higher $h^2_{snp}$ 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 $h^2_{snp}$= 0.0019 in SCZ and $h^2_{snp}$=0.00063 in BD.


**Polygenic dissection of subphenotypes**

Subphenotypes were collected for a subset of patients in 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[7] 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 from comorbidities, clinical information such as rapid cycling and psychotic features as well as additional disease course data such as and 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. We applied a significance cutoff of $p < 0.0004$ based on Bonferroni multiple test correction for 112 tests. In total, we identified 6 significant results after correction (Figure 3, Table 1). For BD PRS we see a significant positive correlation between PRS and manic symptoms in SCZ cases as seen previously[7] ($p=2x10^{-5}$, t=4.26) and psychotic features in BD patients ($p=5.3x10^{-5}$, t=4.04). For SCZ PRS, we see a significant increase in PRS for BD cases with versus without psychotic features ($p=1.2x10^{-10}$, t=6.45) and negative symptoms in SCZ patients ($p=3.60x10^{-6}$, t=4.64). As with the SCZ PRS, BD+SCZ PRS is also 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). While not surpassing conservative correction, the next two most significant results are 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 correlation between SCZ PRS and psychotic features and identified significant correlation when restricting to only BD type I cases (BDI: 3,763 with psychosis, 2,629 without, $p=1.55x10^{-5}$, Supplementary Table 3).

For all 8 quantitative subphenotypes and 9 binary subphenotypes having at least 1,000 cases, we performed a GWAS within cases to calculate heritability and genetic correlation with BD and SCZ. Only two subphenotypes, had significant $h^2_{snp}$ estimates using LD-score regression[33], 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 significant genetic correlation with SCZ ($r_g$=0.34, SE=0.13, p=0.009). While genetic correlation demonstrates a genome-wide relationship between common variants contributing to SCZ and those contributing to psychotic features in BD cases, we sought to assess whether this could be demonstrated among the most significantly associated SCZ loci. Of the 105 autosomal genome-wide significant SCZ loci previously published[16], 60 out of 100 variants in our dataset after QC demonstrated the same direction of effect for 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 samples. As previously shown[34], we found an extensive degree of genetic sharing between these two disorders. We identified 114 genome-wide significant loci contributing to both disorders of which 37 are novel to this analysis. Despite the high degree of sharing, we identified several loci that significantly differentiated between the two disorders, having opposite directions of effect, and 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 is in *DARS2* (coding for the mitochondrial Aspartate-tRNA ligase) which is highly expressed in the brain and significantly regulated by the most significant SNP rs56355601 ($p_{eQTL}$=2.5x10$^{-11}$). 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[35]. Interestingly, 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[36]. The second most significant locus maps to *ARFGEF2*, which codes for ADP Ribosylation Factor Guanine Nucleotide Exchange Factor 2 (also known as BIG2), a protein involved in vesicular trafficking from the trans-Golgi network. Mutations in *ARFGEF2* have been shown to underlie an autosomal recessive condition characterized by microcephaly and periventricular heterotopia, a disorder caused by abnormal neural proliferation and migration[37]. Although not genome-wide significant the third most significant locus implicates *ARNTL* (Aryl Hydrocarbon Receptor Nuclear Translocator Like), which is a core component of the circadian clock and previously hypothesized for relevance in bipolar disorder,[38] although human genetic evidence is limited[39]. Incorporating transcriptional data identified a third genome-wide significant finding in *DCAKD*. The gene codes for Dephospho-CoA Kinase Domain Containing, a member of the human postsynaptic density proteome from human neocortex[40]. In the mouse cortical synaptoproteome *DCAKD* has been found to be among the proteins with the highest changes between juvenile postnatal days and adult stage, which suggests a putative role in brain development[41,42].

We further assessed the contribution of regions of the genome to each disorder through joint regional association and regional heritability estimation. These results point to two additional loci 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* to chromatin, 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 to autism[43]. Their members, including *GATAD2A*, display preferential expression in fetal brain development[43] and in recent work has been implicated to SCZ through open chromatin[44]. Further, p66α (mouse *GATAD2A*) has been recently shown to participate in memory preservation through long-lasting histone modifications in hippocampal memory-activated neurons[45]. The region on chr10 appears to be shared across both disorders, however there are additional independent contributing variants to SCZ and not BD indicating another region of interest, although biological interpretation remains unknown.

More broadly, 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 BD loci being specific to BD. However, we cannot exclude

that the asymmetry of results may less power of discovery for BD than SCZ. The degree in which power and subphenotypes contribute to this result requires further examination.

We have now identified multiple genomic signatures that correlate between one disorder and a clinical symptom in the other disorder, demonstrating that there are genetic components underlying particular symptom dimensions within these disorders. As previously shown, we find a significant positive correlation between 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 BD subtype. Further, we show evidence 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 an independently 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 and identify variants contributing to that symptom dimension. All in all, this work illustrates the utility of genetic data to dissect symptom heterogeneity among correlated disorders and further work could potentially aid in defining sub groups of patients for more personalized treatment.

**References**

1. Whiteford, H. A. *et al.* Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. *The Lancet* **382,** 1575–1586 (2013).

2. Nöthen, M. M., Nieratschker, V., Cichon, S. & Rietschel, M. New findings in the genetics of major psychoses. *Dialogues Clin. Neurosci.* **12,** 85–93 (2010).

3. Lichtenstein, P. *et al.* Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *The Lancet* **373,** 234–239 (2009).

4. Diagnostic and Statistical Manual of Mental Disorders | DSM Library. Available at: http://dsm.psychiatryonline.org/doi/book/10.1176/appi.books.9780890425596. (Accessed: 14th March 2017)

5. WHO | International Classification of Diseases. *WHO* Available at: http://www.who.int/classifications/icd/en/. (Accessed: 14th March 2017)

6. Bortolato, B., Miskowiak, K. W., Köhler, C. A., Vieta, E. & Carvalho, A. F. Cognitive dysfunction in bipolar disorder and schizophrenia: a systematic review of meta-analyses. *Neuropsychiatr. Dis. Treat.* **11,** 3111–3125 (2015).

7. Ruderfer, D. M. *et al.* Polygenic dissection of diagnosis and clinical dimensions of bipolar disorder and schizophrenia. *Mol. Psychiatry* **19,** 1017–1024 (2014).

8. Kirov, G. *et al.* De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Mol. Psychiatry* **17,** 142–153 (2012).

9. Szatkiewicz, J. P. *et al.* Copy number variation in schizophrenia in Sweden. *Mol. Psychiatry* **19,** 762–773 (2014).

10.     CNV and Schizophrenia Working Groups of the Psychiatric Genomics Consortium. Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nat. Genet.* **49,** 27–35 (2017).

11.     Stone, J. L. *et al.* Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* **455,** 237–241 (2008).

12.     Gulsuner, S. & McClellan, J. M. Copy Number Variation in Schizophrenia. *Neuropsychopharmacology* **40,** 252–254 (2015).

13.     Green, E. K. *et al.* Copy number variation in bipolar disorder. *Mol. Psychiatry* **21,** 89–93 (2016).

14.     Kataoka, M. *et al.* Exome sequencing for bipolar disorder points to roles of de novo loss-of-function and protein-altering mutations. *Mol. Psychiatry* **21,** 885–893 (2016).

15.     Fromer, M. *et al.* De novo mutations in schizophrenia implicate synaptic networks. *Nature* **506,** 179–184 (2014).

16.     Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511,** 421–427 (2014).

17.     Howie, B., Marchini, J. & Stephens, M. Genotype Imputation with Thousands of Genomes. *G3 GenesGenomesGenetics* **1,** 457–470 (2011).

18.     Delaneau, O., Zagury, J.-F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods* **10,** 5–6 (2013).

19.     Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **81,** 559–575 (2007).

20.     Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48,** 481–487 (2016).

21.    Westra, H.-J. *et al.* Systematic identification of trans eQTLs as putative drivers of known

       disease associations. *Nat. Genet.* **45,** 1238–1243 (2013).

22.    Consortium, T. Gte. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue

       gene regulation in humans. *Science* **348,** 648–660 (2015).

23.    Pickrell, J. K. *et al.* Detection and interpretation of shared genetic influences on 42

       human traits. *Nat. Genet.* **48,** 709–717 (2016).

24.    Berisa, T. & Pickrell, J. K. Approximately independent linkage disequilibrium blocks in

       human populations. *Bioinformatics* **32,** 283–285 (2015).

25.    Shi, H., Kichaev, G. & Pasaniuc, B. Contrasting the Genetic Architecture of 30 Complex

       Traits from Summary Association Data. *Am. J. Hum. Genet.* **99,** 139–153 (2016).

26.    The 1000 Genomes Project Consortium. A global reference for human genetic variation.

       *Nature* **526,** 68–74 (2015).

27.    Giambartolomei, C. *et al.* Bayesian Test for Colocalisation between Pairs of Genetic

       Association Studies Using Summary Statistics. *PLOS Genet.* **10,** e1004383 (2014).

28.    He, X. *et al.* Sherlock: Detecting Gene-Disease Associations by Matching Patterns of

       Expression QTL and GWAS. *Am. J. Hum. Genet.* **92,** 667–680 (2013).

29.    Gamazon, E. R. *et al.* A gene-based association method for mapping traits using

       reference transcriptome data. *Nat. Genet.* **47,** 1091–1098 (2015).

30.    Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association

       studies. *Nat. Genet.* **48,** 245–252 (2016).

31.    Fromer, M. *et al.* Gene expression elucidates functional impact of polygenic risk for

       schizophrenia. *Nat. Neurosci.* **19,** 1442–1453 (2016).

32.     Benner, C. *et al.* FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* **32,** 1493–1501 (2016).

33.     Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47,** 291–295 (2015).

34.     Cross-Disorder Group of the Psychiatric Genomics Consortium. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* **45,** 984–994 (2013).

35.     Yamashita, S. *et al.* Neuropathology of leukoencephalopathy with brainstem and spinal cord involvement and high lactate caused by a homozygous mutation of DARS2. *Brain Dev.* **35,** 312–316 (2013).

36.     Luoni, A. *et al.* Ankyrin-3 as a molecular marker of early-life stress and vulnerability to psychiatric disorders. *Transl. Psychiatry* **6,** e943 (2016).

37.     Sheen, V. L. *et al.* Mutations in ARFGEF2 implicate vesicle trafficking in neural progenitor proliferation and migration in the human cerebral cortex. *Nat. Genet.* **36,** 69–76 (2004).

38.     Yang, S., Van Dongen, H. P. A., Wang, K., Berrettini, W. & Bućan, M. Assessment of circadian function in fibroblasts of patients with bipolar disorder. *Mol. Psychiatry* **14,** 143–155 (2008).

39.     Byrne, E. M. *et al.* Testing the role of circadian genes in conferring risk for psychiatric disorders. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet. Off. Publ. Int. Soc. Psychiatr. Genet.* **0,** 254–260 (2014).

40.     Bayés, À. *et al.* Characterization of the proteome, diseases and evolution of the human postsynaptic density. *Nat. Neurosci.* **14,** 19–21 (2011).

41.     Moczulska, K. E. *et al.* Deep and Precise Quantification of the Mouse Synaptosomal Proteome Reveals Substantial Remodeling during Postnatal Maturation. *J. Proteome Res.* **13,** 4310–4324 (2014).

42.     Gonzalez-Lozano, M. A. *et al.* Dynamics of the mouse brain cortical synaptic proteome during postnatal brain development. *Sci. Rep.* **6,** (2016).

43.     Li, J. *et al.* Identification of Human Neuronal Protein Complexes Reveals Biochemical Activities and Convergent Mechanisms of Action in Autism Spectrum Disorders. *Cell Syst.* **1,** 361–374 (2015).

44.     Fullard, J. F. *et al.* Open chromatin profiling of human postmortem brain infers functional roles for non-coding schizophrenia loci. *Hum. Mol. Genet.* doi:10.1093/hmg/ddx103

45.     Ding, X. *et al.* Activity-induced histone modifications govern Neurexin-1 mRNA splicing and memory preservation. *Nat. Neurosci.* **advance online publication,** (2017).
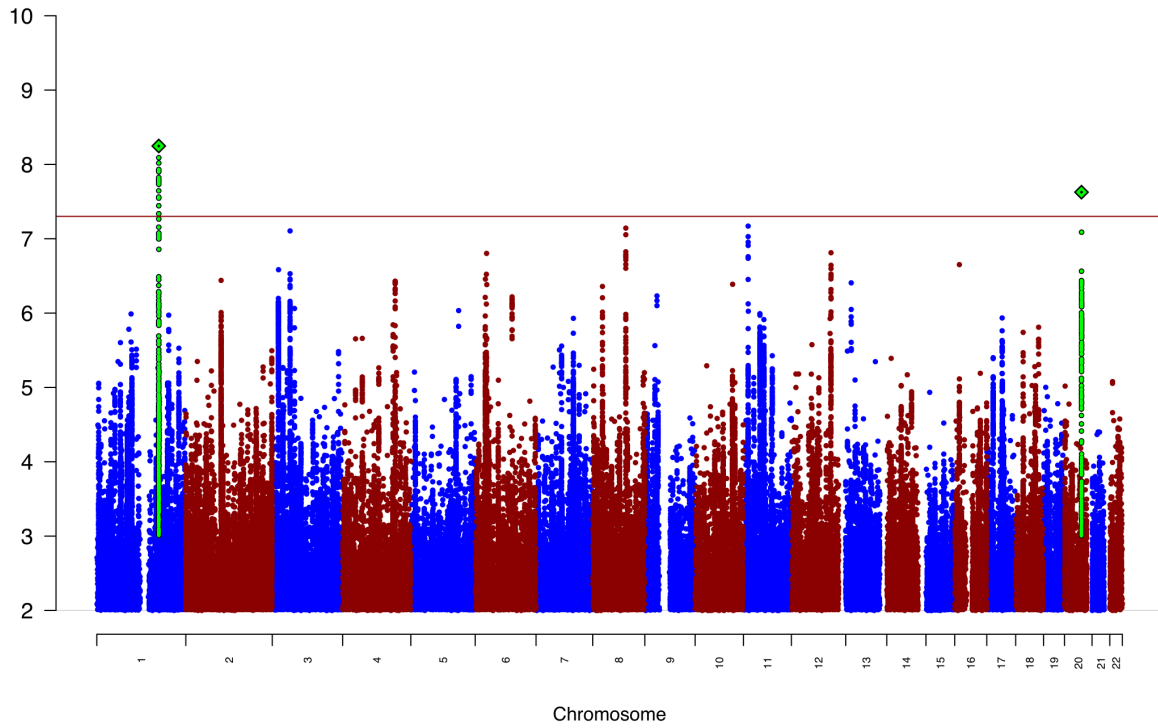
# Figures

**a.**
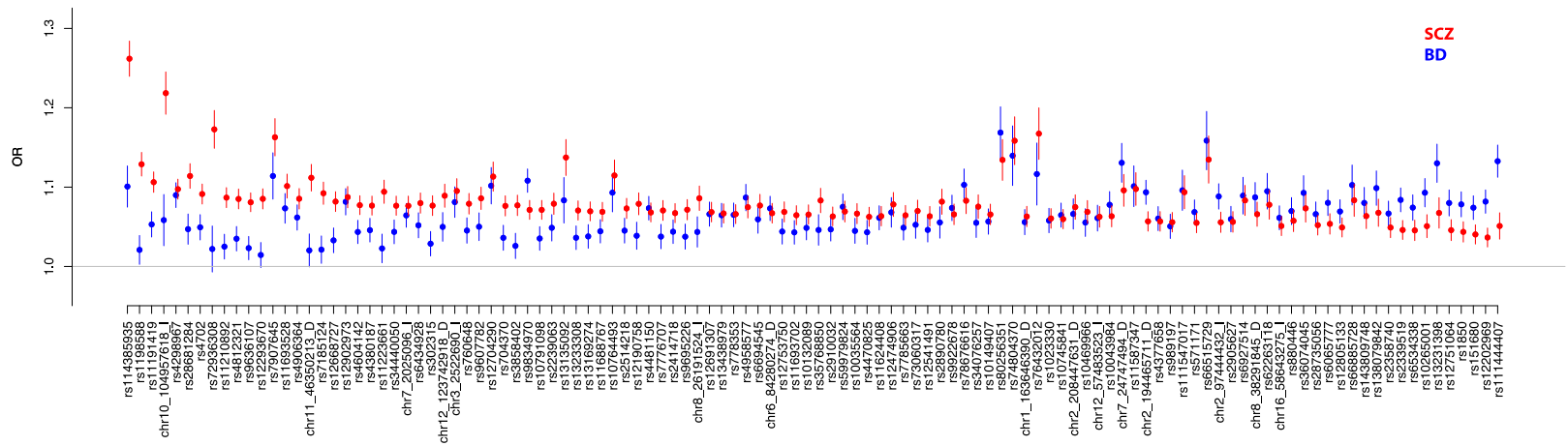


**b.**

**Figure 1.** 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.
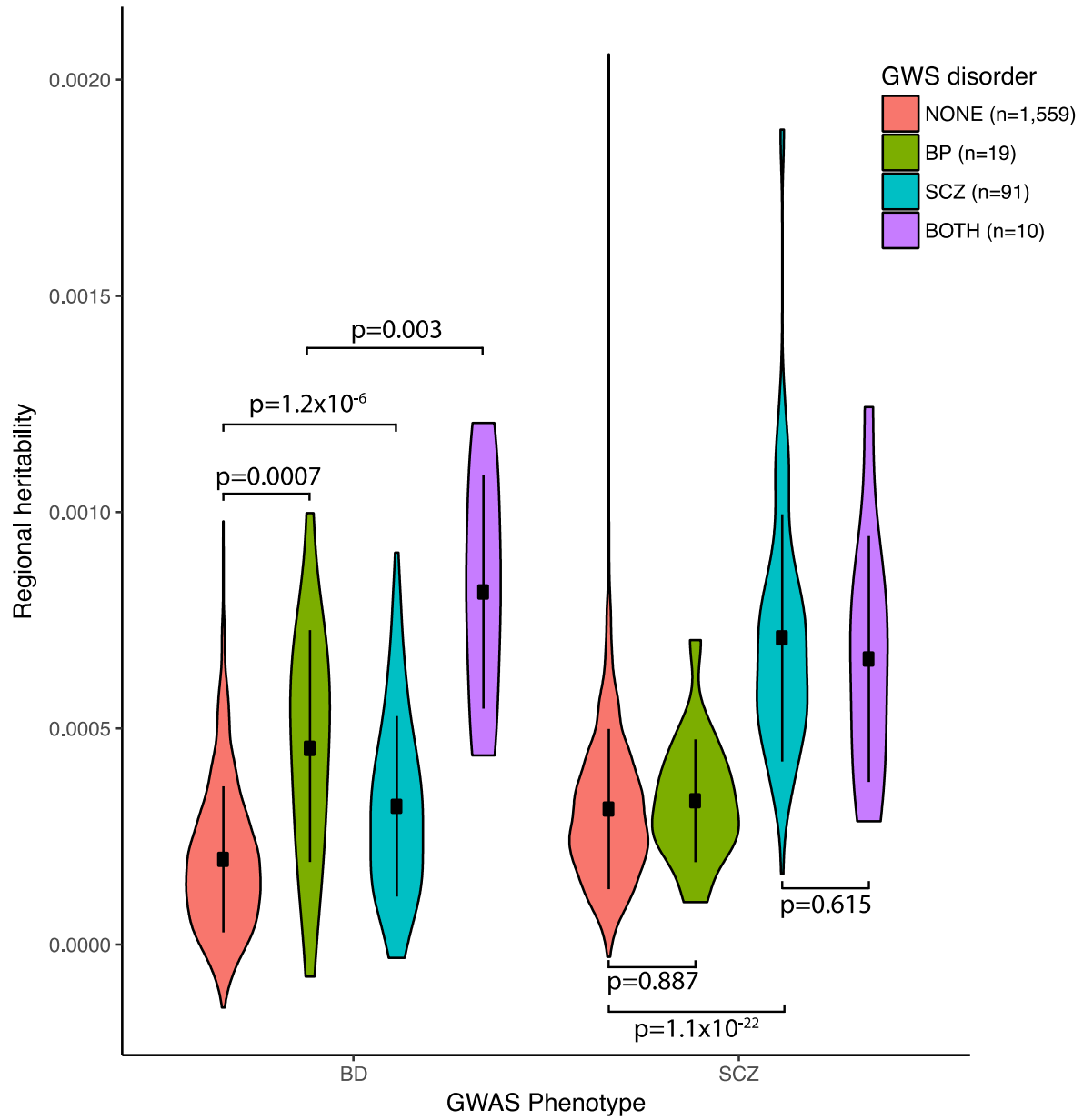
**Figure 2.** Regional SNP-heritability estimates for SCZ and BD stratified by whether the region contains the most significant variant in a genome-wide significant locus in BD, SCZ, neither or both.
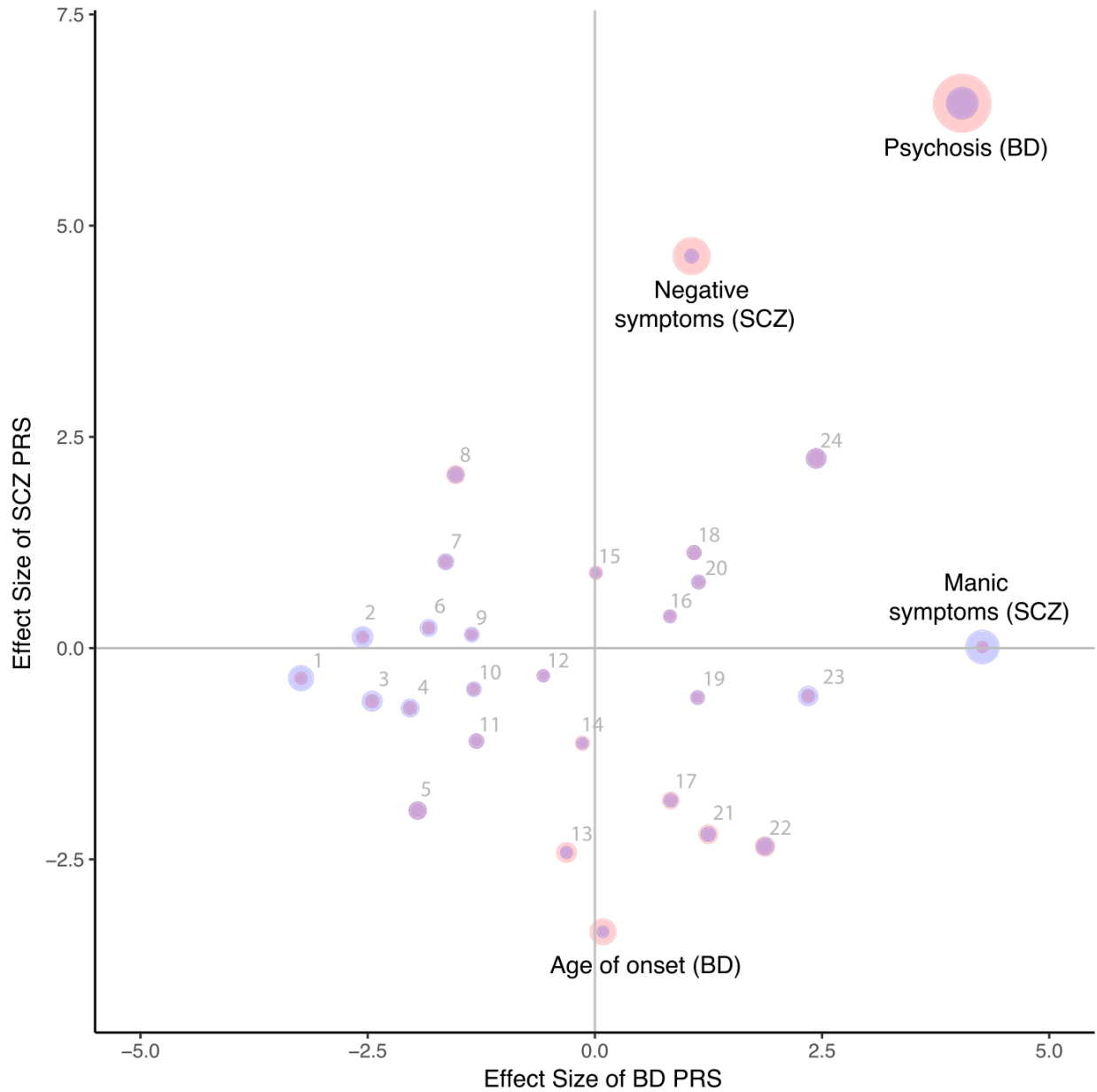
**Figure 3.** Effect size (calculated by dividing regression estimate by standard error) from regression analysis including ancestry covariates for each subphenotype and PRS for BD (x-axis) and SCZ (y-axis). Point size represents –log10(p-value) with SCZ (red) and BD (blue). Numbered subphenotypes are 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 1.** 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 ancestry and site as covariates.

Effect is the regression estimate divided by the standard error.

| | Subphenotype | N | Cases | Controls | *P-value* | | | | *Effect* | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | BP+SCZ | BP | SCZ | SCZvsBD | BP+SCZ | BP | SCZ | SCZvsBD |
| BD | psychosis | 8131 | 4632 | 3499 | **7.9E-13** | **5.3E-05** | **1.2E-10** | 5.8E-01 | 7.17 | 4.04 | 6.45 | 0.55 |
| | suicide ideation | 5399 | 3801 | 1598 | 7.8E-01 | 1.8E-01 | 8.7E-01 | 1.7E-01 | -0.28 | -1.35 | 0.16 | 1.37 |
| | family history | 4971 | 2730 | 2241 | 6.1E-02 | 2.8E-01 | 2.6E-01 | 6.9E-01 | 1.87 | 1.09 | 1.13 | -0.39 |
| | irritable mania | 4230 | 2401 | 1829 | 3.8E-01 | 4.1E-01 | 7.1E-01 | 1.0E-01 | 0.88 | 0.83 | 0.38 | -1.63 |
| | rapid cycling | 5214 | 1744 | 3470 | 7.9E-03 | 5.1E-02 | 5.5E-02 | 3.1E-01 | -2.66 | -1.95 | -1.92 | 1.01 |
| | alcohol substance dependence | 5440 | 1494 | 3946 | 4.5E-01 | 2.1E-01 | 2.8E-02 | 1.7E-01 | -0.75 | 1.25 | -2.20 | -1.36 |
| | panic disorder | 4647 | 863 | 3784 | 2.8E-01 | 1.8E-01 | 6.3E-01 | 4.0E-01 | -1.07 | -1.33 | -0.49 | 0.83 |
| | panic attacks | 3976 | 851 | 3125 | 1.3E-01 | 1.1E-02 | 9.0E-01 | 4.7E-02 | -1.50 | -2.56 | 0.13 | 1.98 |
| | mixed states | 4044 | 826 | 3218 | 1.0E-01 | 4.2E-02 | 4.8E-01 | 6.0E-02 | -1.64 | -2.03 | -0.71 | 1.88 |
| | unipolar mania | 4863 | 461 | 4402 | 2.4E-02 | 2.5E-01 | 4.3E-01 | 6.1E-01 | 2.26 | 1.14 | 0.78 | 0.51 |
| | comorbid migraine | 2652 | 410 | 2242 | 1.3E-02 | 1.2E-03 | 7.2E-01 | 4.4E-01 | -2.48 | -3.23 | -0.36 | 0.77 |
| | comorbid OCD | 4215 | 386 | 3829 | 9.7E-01 | 1.0E-01 | 3.1E-01 | 1.9E-01 | -0.04 | -1.64 | 1.02 | 1.30 |
| | comorbid eating disorder | 3839 | 331 | 3508 | 2.1E-01 | 6.7E-02 | 8.1E-01 | 6.3E-01 | -1.25 | -1.83 | 0.24 | 0.48 |
| | age of onset | 8610 | | | 6.2E-03 | 9.3E-01 | 7.9E-04 | 6.2E-01 | -2.74 | 0.09 | -3.36 | -0.50 |
| | age at interview | 8062 | | | 5.9E-01 | 1.9E-02 | 5.7E-01 | 4.4E-01 | 0.54 | 2.35 | -0.57 | -0.78 |
| | episodes mixed mania | 6587 | | | 6.3E-01 | 2.6E-01 | 5.6E-01 | 3.2E-01 | -0.48 | 1.13 | -0.58 | -1.00 |
| | suicide attempt | 6308 | | | 1.2E-01 | 1.4E-02 | 5.3E-01 | 2.8E-01 | -1.54 | -2.45 | -0.63 | 1.09 |
| | episodes depressive | 6252 | | | 7.4E-03 | 7.6E-01 | 1.6E-02 | 9.6E-01 | -2.68 | -0.31 | -2.42 | -0.05 |
| | episodes total | 5958 | | | 1.3E-01 | 8.9E-01 | 2.6E-01 | 3.9E-01 | -1.51 | -0.14 | -1.13 | -0.87 |
| | year of birth | 5317 | | | 1.7E-01 | 1.3E-01 | 4.0E-02 | 3.6E-02 | 1.39 | -1.53 | 2.05 | 2.10 |
| | number of suicide attempts | 5015 | | | 6.2E-02 | 1.9E-01 | 2.7E-01 | 4.9E-01 | -1.87 | -1.30 | -1.10 | -0.69 |
| | number of hospitalizations | 3944 | | | 4.2E-04 | 1.5E-02 | 2.5E-02 | 7.4E-01 | 3.53 | 2.43 | 2.25 | -0.33 |
| | age of onset depression | 3467 | | | 2.3E-01 | 4.0E-01 | 7.2E-02 | 2.2E-01 | -1.19 | 0.83 | -1.80 | 1.24 |
| | age of onset mania | 3395 | | | 2.5E-01 | 6.1E-02 | 1.9E-02 | 2.2E-01 | -1.14 | 1.87 | -2.35 | -1.23 |
| SCZ | Manic | 6908 | | | 2.4E-02 | **2.0E-05** | 9.9E-01 | 3.5E-02 | 2.26 | 4.26 | 0.01 | -2.10 |
| | Depressive | 6907 | | | 9.0E-01 | 5.7E-01 | 7.4E-01 | 1.8E-01 | 0.13 | -0.57 | -0.33 | -1.36 |
| | Negative | 8355 | | | **1.5E-05** | 2.9E-01 | **3.6E-06** | 2.1E-02 | 4.33 | 1.06 | 4.64 | 2.31 |
| | Positive | 8259 | | | 4.1E-01 | 9.9E-01 | 3.7E-01 | 5.1E-01 | 0.82 | 0.01 | 0.89 | 0.65 |