Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology
HUNT All-In Psychiatry

**published in**
Nature genetics
2021

**DOI (link to publisher)**
10.1038/s41588-021-00857-4

**document version**
Publisher's PDF, also known as Version of record

**document license**
Article 25fa Dutch Copyright Act

**Link to publication in VU Research Portal**

**citation for published version (APA)**
HUNT All-In Psychiatry (2021). Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nature genetics*, 53(6), 817-829. https://doi.org/10.1038/s41588-021-00857-4

**General rights**
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal?

**Take down policy**
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

**E-mail address:**
vuresearchportal.ub@vu.nl

Download date: 16. Sep. 2023
Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology

Bipolar disorder is a heritable mental illness with complex etiology. We performed a genome-wide association study of 41,917 bipolar disorder cases and 371,549 controls of European ancestry, which identified 64 associated genomic loci. Bipolar disorder risk alleles were enriched in genes in synaptic signaling pathways and brain-expressed genes, particularly those with high specificity of expression in neurons of the prefrontal cortex and hippocampus. Significant signal enrichment was found in genes encoding targets of antipsychotics, calcium channel blockers, antiepileptics and anesthetics. Integrating expression quantitative trait locus data implicated 15 genes robustly linked to bipolar disorder via gene expression, encoding druggable targets such as HTR6, MCHR1, DCLK3 and FURIN. Analyses of bipolar disorder subtypes indicated high but imperfect genetic correlation between bipolar disorder type I and II and identified additional associated loci. Together, these results advance our understanding of the biological etiology of bipolar disorder, identify novel therapeutic leads and prioritize genes for functional follow-up studies.

Bipolar disorder (BD) is a complex mental disorder characterized by recurrent episodes of (hypo)mania and depression. It is a common condition affecting an estimated 40 to 50 million people worldwide. This, combined with the typical onset in young adulthood, an often chronic course and increased risk of suicide, makes BD a major public health concern and a major cause of global disability. Clinically, BD is classified into two main subtypes: bipolar I disorder (BD I), in which manic episodes typically alternate with depressive episodes, and bipolar II disorder (BD II), characterized by the occurrence of at least one hypomanic and one depressive episode. These subtypes have a lifetime prevalence of ~1% each in the population.

Family and molecular genetic studies provide convincing evidence that BD is a multifactorial disorder, with genetic and environmental factors contributing to its development. On the basis of twin and family studies, the heritability of BD is estimated at 60–85%8–10. Genome-wide association studies (GWASs)11–20 have led to valuable insights into the genetic etiology of BD. The largest such study has been conducted by the Psychiatric Genomics Consortium (PGC), in which genome-wide SNP data from 29,764 patients with BD and 169,118 controls were analyzed and 30 genome-wide significant loci were identified (PGC2)15. SNP-based heritability ($h_{SNP}^2$) estimation using the same data suggested that common genetic variants genome-wide explain ~20% of BD’s phenotypic variance16. Polygenic risk scores (PRSs) generated from the results of this study explained ~4% of phenotypic variance in independent samples. Across the genome, genetic associations with BD converged on specific biological pathways including regulation of insulin secretion17–19, retrograde endocannabinoid signaling20, glutamate receptor signaling21 and calcium channel activity22.

Despite this considerable progress, only a fraction of the genetic etiology of BD has been identified, and the specific biological mechanisms underlying the development of the disorder are still unknown. In the present study, we report the results of the third GWAS meta-analysis of the PGC Bipolar Disorder Working Group, comprising 41,917 individuals with BD and 371,549 controls. These results confirm and expand on many previously reported findings, identify novel therapeutic leads and prioritize genes for functional follow-up studies23,24. Thus, our results further illuminate the biological etiology of BD.

Results

GWAS results. We conducted a GWAS meta-analysis of 57 BD cohorts collected in Europe, North America and Australia (Supplementary Table 1), totaling 41,917 individuals with BD (cases) and 371,549 controls of European descent (effective $n = 101,962$, see Methods). For 52 cohorts, individual-level genotype and phenotype data were shared with the PGC and cases met international consensus criteria (DSM-IV, ICD-9 or ICD-10) for lifetime BD, established using structured diagnostic interviews, clinician-administered checklists or medical record review. BD GWAS summary statistics were received for five external cohorts (iPSYCH25, deCODE genetics26, Estonian Biobank27, Trøndelag Health Study (HUNT)28 and UK Biobank29), in which most cases were ascertained using ICD codes. The GWAS meta-analysis identified 64 independent loci associated with BD at genome-wide significance ($P < 5 \times 10^{-8}$; Fig. 1, Table 1 and Supplementary Table 2). Using linkage disequilibrium score regression (LDSC)30, the $h_{SNP}^2$ of BD was estimated to be 18.6% (s.e. = 0.008, $P = 5.1 \times 10^{-129}$) on the liability scale, assuming a BD population prevalence of 2%, and 15.6% (s.e. = 0.006, $P = 5.0 \times 10^{-132}$) assuming a population prevalence of 1% (Supplementary Table 3). The genomic inflation factor ($\lambda_{GC}$) was 1.38 and the LDSC intercept was 1.04 (s.e. = 0.01, $P = 2.5 \times 10^{-4}$; Supplementary Fig. 1). While the intercept has frequently been used as an indicator of confounding from population stratification, it can rise above 1 with increased sample size and heritability. The attenuation ratio—(LDSC intercept − 1)/(mean of association chi-square)—which is not subject to these limitations, was 0.06 (s.e. = 0.02), indicating that the majority of inflation of the GWAS test statistics was due to polygenicity31,32. Of the 64 genome-wide significant loci, 33 are novel discoveries (that is, loci not overlapping with any locus previously reported as genome-wide significant for BD). Novel loci include the major histocompatibility complex (MHC) and loci previously reaching genome-wide significance for other psychiatric disorders, including ten for schizophrenia, four for major depression and three for childhood-onset psychiatric disorders or problematic alcohol use (Table 1).

A full list of authors and their affiliations appears at the end of the paper.

NATURE GENETICS | VOL 53 | JUNE 2021 | 817–829 | www.nature.com/naturegenetics
Enrichment analyses. Genome-wide analyses using MAGMA indicated significant enrichment of BD associations in 161 genes (Supplementary Table 4) and 4 gene sets related to synaptic signaling (Supplementary Table 5). The BD association signal was enriched among genes expressed in different brain tissues (Supplementary Table 6), especially genes with high specificity of gene expression in neurons (both excitatory and inhibitory) versus other cell types, within cortical and subcortical brain regions in mice (Supplementary Fig. 2). In human brain samples, signal enrichment was also observed in hippocampal pyramidal neurons and interneurons of the prefrontal cortex and hippocampus, compared with other cell types (Supplementary Fig. 2).

In a gene-set analysis of the targets of individual drugs (from the Drug–Gene Interaction Database DGIdb 2.0 [ref. 19] and the Psychoactive Drug Screening Database Ki DB [40]), the targets of the calcium channel blockers mibefradil and nisoldipine were significantly enriched (Supplementary Table 7). Grouping drugs according to their anatomical therapeutic chemical classes [41], there was significant enrichment in the targets of four broad drug classes (Supplementary Table 8): psycholeptics (drugs with a calming effect on behavior; especially hypnotics and sedatives, antipsychotics and anxiolytics), calcium channel blockers, antiepileptics and (general) anesthetics (Supplementary Table 8).

Expression quantitative trait locus integrative analyses. We conducted a transcriptome-wide association study (TWAS) using FUSION and expression quantitative trait locus (eQTL) data from the PsychENCODE Consortium (1,321 brain samples) [43]. BD-associated alleles significantly influenced expression of 77 genes in the brain (Supplementary Table 9 and Supplementary Fig. 3). These genes encompassed 40 distinct regions. We performed TWAS fine-mapping using FOCUS to model the correlation among the TWAS signals and prioritize the most likely causal gene(s) in each region. Within the 90%-credible set, FOCUS prioritized 22 genes with a posterior inclusion probability (PIP) > 0.9 (encompassing 20 distinct regions) and 32 genes with a PIP > 0.7 (29 distinct regions; Supplementary Table 10).

We used summary-data-based Mendelian randomization (SMR) to identify putative causal relationships between SNPs and BD via gene expression by integrating the BD GWAS results with brain eQTL summary statistics from the PsychENCODE Consortium and blood eQTL summary statistics from the eQTLGen consortium (31,684 whole blood samples). The eQTLGen results represent the largest existing eQTL study and provide independent eQTL data. Of the 32 genes fine-mapped with PIP > 0.7, 15 were significantly associated with BD in the SMR analyses and passed the heterogeneity in dependent instruments (HEIDI) test, suggesting that their effect on BD is mediated via gene expression in the brain and/or blood (Supplementary Table 11). The genes located in genome-wide significant loci are labeled in Fig. 1. Other significant genes included HTR6, DCLK3, HAPLN4 and PACSIN2.

MHC locus. Variants within and distal to the MHC locus were associated with BD at genome-wide significance. The most highly associated SNP was rs13195402, 3.2 megabases (Mb) distal to any HLA gene or the complement component 4 (C4) genes (Supplementary Fig. 4). Imputation of C4 alleles using SNP data uncovered no association between the five most common structural forms of the C4A/C4B locus (BS, AL–BS, AL–BL and AL–AL) and BD, either before or after conditioning on rs13195402 (Supplementary Fig. 5). While genetically predicted C4A expression initially showed a weak association with BD, this association was nonsignificant after conditioning on rs13195402 (Supplementary Fig. 5).

Polygenic risk scoring. The performance of PRSs based on these GWAS results was assessed by excluding cohorts in turn from the meta-analysis to create independent test samples. PRSs explained ~4.5% of phenotypic variance in BD on the liability scale (at GWAS P-value threshold (GWAS P < 0.1, BD population prevalence 2%), based on the weighted mean R² across cohorts (Fig. 2).
| Locus | CHR | BP | SNP | P     | OR  | s.e. | A1/A2 | A1 freq in controls | Previous report* for BD (citation) | Name for novel locus* | Previous report* for psychiatric disorders |
|-------|-----|-----|-----|-------|-----|------|-------|---------------------|-----------------------------------|----------------------|------------------------------------------|
| 1     | 1   | 6105668 | rs2126180 | 1.6 × 10⁻⁵ | 1.058 | 0.009 | A/G | 0.457 | LINC01748 |
| 2     | 1   | 163475389 | rs10737496 | 7.2 × 10⁻⁵ | 1.056 | 0.009 | C/T | 0.444 | NUF2 |
| 3     | 1   | 97416153 | rs4619651 | 4.8 × 10⁻⁴ | 1.068 | 0.010 | G/A | 0.670 | LMAN2L (PGC2) |
| 4     | 2   | 166152389 | rs17183814 | 2.7 × 10⁻⁸ | 1.108 | 0.019 | G/A | 0.924 | SCN2A (PGC2) |
| 5     | 2   | 169481837 | rs13417268 | 2.1 × 10⁻⁵ | 1.064 | 0.011 | C/G | 0.758 | CERS6 |
| 6     | 2   | 193738336 | rs2011302 | 4.3 × 10⁻⁸ | 1.055 | 0.010 | A/T | 0.377 | |
| 7     | 2   | 194437889 | rs2719164 | 4.9 × 10⁻⁴ | 1.053 | 0.010 | A/G | 0.564 | Intergenic (PGC2) |
| 8     | 3   | 36856030 | rs9834970 | 6.6 × 10⁻⁹ | 1.087 | 0.009 | C/T | 0.481 | TRANK1 (PGC2) |
| 9     | 3   | 52626443 | rs2336147 | 3.6 × 10⁻³ | 1.070 | 0.009 | T/C | 0.498 | ITIH1 (PGC2) |
| 10    | 3   | 70488788 | rs115694474 | 2.4 × 10⁻⁸ | 1.068 | 0.012 | T/A | 0.799 | MDFC2 |
| 11    | 3   | 107757060 | rs696366 | 4.5 × 10⁻⁴ | 1.053 | 0.009 | C/A | 0.550 | CD47 (PGC2) |
| 12    | 4   | 123076007 | rs112481526 | 1.9 × 10⁻⁸ | 1.065 | 0.011 | G/A | 0.256 | KIA1109 |
| 13    | 5   | 7542991 | rs28565152 | 2.0 × 10⁻⁹ | 1.070 | 0.011 | A/G | 0.238 | ADCY2 (PGC2) |
| 14    | 5   | 78849505 | rs6865469 | 1.7 × 10⁻⁵ | 1.060 | 0.010 | T/G | 0.274 | HOMER1 |
| 15    | 5   | 80961069 | rs6887473 | 8.8 × 10⁻⁸ | 1.062 | 0.011 | G/A | 0.739 | SSBP2 (PGC2) |
| 16    | 5   | 13771212 | rs10043984 | 3.7 × 10⁻⁸ | 1.062 | 0.011 | T/C | 0.236 | KDM3B |
| 17    | 5   | 169289206 | rs10866641 | 2.8 × 10⁻¹⁰ | 1.065 | 0.009 | T/C | 0.575 | DOCK2 |
| 18    | 6   | 26463575 | rs13195402 | 5.8 × 10⁻⁸ | 1.146 | 0.018 | G/T | 0.919 | MHC |
| 19    | 6   | 98565211 | rs1487445 | 1.5 × 10⁻⁵ | 1.078 | 0.009 | T/C | 0.487 | POU3F2 (PGC2) |
| 20    | 6   | 152793572 | rs431993 | 2.0 × 10⁻⁸ | 1.056 | 0.010 | A/T | 0.382 | SYNE1 (Green et al., 2013) |
| 21    | 6   | 166995260 | rs10459979 | 4.2 × 10⁻⁸ | 1.057 | 0.010 | G/C | 0.500 | RPS6KA2 (PGC2) |
| 22    | 7   | 2020995 | rs12668848 | 1.9 × 10⁻⁹ | 1.059 | 0.010 | G/A | 0.575 | MADIL1 (Hou et al., 2016, Ikeda et al., 2018) |
| 23    | 7   | 11871878 | rs13779084 | 1.4 × 10⁻¹³ | 1.079 | 0.010 | A/G | 0.299 | THSD7A (PGC2) |
| 24    | 7   | 21492589 | rs6954854 | 5.9 × 10⁻¹⁰ | 1.060 | 0.009 | G/A | 0.425 | SP4 |
| 25    | 7   | 24647222 | rs12672003 | 2.7 × 10⁻⁹ | 1.096 | 0.016 | G/A | 0.113 | MPP6 |
| 26    | 7   | 105043229 | rs11764361 | 3.5 × 10⁻⁹ | 1.063 | 0.010 | A/G | 0.668 | SRPK2 (PGC2) |
| 27    | 7   | 131870597 | rs6946056 | 3.7 × 10⁻⁸ | 1.055 | 0.010 | C/A | 0.623 | PLXNA4 |
| 28    | 7   | 140676153 | rs10255167 | 1.6 × 10⁻⁶ | 1.068 | 0.012 | A/G | 0.778 | MRPS33 (PGC2) |
| 29    | 8   | 9763581 | rs62489493 | 2.6 × 10⁻¹¹ | 1.094 | 0.014 | G/C | 0.128 | miR124-1 |
| 30    | 8   | 10226355 | rs3088186 | 2.1 × 10⁻⁷ | 1.058 | 0.010 | T/C | 0.287 | MSRA |
| 31    | 8   | 34152492 | rs2953928 | 6.3 × 10⁻⁹ | 1.124 | 0.020 | A/G | 0.067 | RPI1-B40J5.2 (lincRNA) |
| 32    | 8   | 144993377 | rs6992333 | 1.6 × 10⁻⁹ | 1.062 | 0.010 | G/A | 0.410 | PLEC |
| 33    | 9   | 37090538 | rs10973201 | 2.5 × 10⁻⁸ | 1.101 | 0.017 | C/T | 0.110 | ZCCHC7 |
| 34    | 9   | 141066490 | rs62581014 | 2.8 × 10⁻⁸ | 1.067 | 0.012 | T/C | 0.366 | TUBB5 |
| 35    | 10  | 1875103 | rs1998820 | 4.1 × 10⁻⁴ | 1.087 | 0.015 | T/A | 0.886 | CACNB2 |
| 36    | 10  | 62322034 | rs10994415 | 1.1 × 10⁻¹¹ | 1.125 | 0.017 | C/T | 0.082 | ANK3 (PGC2) |
| 37    | 10  | 64525135 | rs10761661 | 4.7 × 10⁻⁸ | 1.053 | 0.009 | T/C | 0.472 | ADO |

Continued
and Supplementary Table 12). This corresponds to a weighted mean area under the curve of 65%. Results per cohort and per wave of recruitment to the PGC are in Supplementary Tables 12 and 13 and Supplementary Fig. 7. At GWAS $P < 0.1$, individuals in the top 10% of BD PRSs had an odds ratio of 3.5 (95% confidence interval CI 1.7–7.3) of being affected with the disorder compared with individuals in the middle decile (based on the weighted mean OR across PGC cohorts), and an odds ratio of 9.3 (95% CI 1.7–49.3) compared with individuals in the lowest decile. The generalizability of PRSs from this meta-analysis was examined in several non-European cohorts. PRSs explained up to 2.3% and 1.9% of variance in BD in two East Asian samples, and 1.2% and 0.4% in two admixed African American samples (Fig. 2 and Supplementary Table 14).

The variance explained by the PRSs increased in every cohort with increasing sample size of the PGC BD European discovery sample (Supplementary Fig. 8 and Supplementary Table 14).

**Genetic architecture of BD and other traits.** The genome-wide genetic correlation ($r_g$) of BD with a range of diseases and traits was assessed on LD Hub94. After correction for multiple testing, BD showed significant $r_g$ with 16 traits among 255 tested from published GWASs (Supplementary Table 15). Genetic correlation was positive with all psychiatric disorders assessed, particularly schizophrenia ($r_g = 0.68$) and major depression ($r_g = 0.44$), and to a lesser degree anorexia, attention deficit/hyperactivity disorder and autism spectrum disorder ($r_g \approx 0.2$). We found evidence of positive $r_g$ between BD

### Table 1 | Genome-wide significant loci for BD from meta-analysis of 41,917 cases and 371,549 controls (Continued)

| Locus | CHR | BP | SNP | $P$ | OR | s.e. | A1/ A2 | A1 freq in controls | Previous report\* for BD (citation) | Name for novel locus\* | Previous report\* for psychiatric disorders |
|-------|-----|----|-----|-----|----|------|--------|-------------------|-----------------------------------|---------------------|--------------------------------------|
| 38c  | 10  | 111648659 | rs2273738 | $1.6 \times 10^{-11}$ | 1.096 | 0.014 | T/C | 0.135 | ADD3 (Charney et al. 2017, PGC2) |  |
| 39a  | 11  | 61618608 | rs174592 | $9.9 \times 10^{-5}$ | 1.074 | 0.010 | G/A | 0.395 | FADS2 (PGC2) |  |
| 40   | 11  | 64009879 | rs4672 | $3.4 \times 10^{-9}$ | 1.107 | 0.017 | A/G | 0.083 |  |  |
| 41c  | 11  | 65848738 | rs475805 | $2.0 \times 10^{-9}$ | 1.070 | 0.011 | A/G | 0.767 | PACS1 (PGC2) |  |
| 42c  | 11  | 66324583 | rs678397 | $5.5 \times 10^{-9}$ | 1.056 | 0.009 | T/C | 0.457 | PC (PGC1, PGC2) |  |
| 43c  | 11  | 705175675 | rs12575685 | $1.2 \times 10^{-10}$ | 1.067 | 0.010 | A/G | 0.327 | SHANK2 (PGC2) |  |
| 44   | 11  | 79092527 | rs12289486 | $3.3 \times 10^{-8}$ | 1.086 | 0.015 | T/C | 0.115 | ODZ4 (PGC1) |  |
| 45c  | 12  | 2348844 | rs11062170 | $1.9 \times 10^{-15}$ | 1.081 | 0.010 | C/G | 0.333 | CACNAIC (PGC2) |  |
| 46   | 13  | 113869045 | rs35306827 | $3.6 \times 10^{-9}$ | 1.068 | 0.011 | G/A | 0.775 | CUL4A |  |
| 47   | 14  | 9971929 | rs2693698 | $2.0 \times 10^{-8}$ | 1.055 | 0.009 | G/A | 0.551 | BCL11B | SCZ, CDG |
| 48c  | 15  | 38973793 | rs35958438 | $3.8 \times 10^{-8}$ | 1.066 | 0.012 | G/A | 0.772 | C1orf53 | CDG |
| 49c  | 15  | 42904904 | rs4447398 | $2.6 \times 10^{-9}$ | 1.086 | 0.014 | A/C | 0.131 | STARD9 (PGC2) |  |
| 50   | 15  | 83531774 | rs62011709 | $1.4 \times 10^{-8}$ | 1.064 | 0.011 | T/A | 0.747 | HOMER2 | SCZ |
| 51c  | 15  | 85149575 | rs748455 | $5.0 \times 10^{-11}$ | 1.070 | 0.010 | T/C | 0.719 | ZNF592 (PGC2) |  |
| 52   | 15  | 91426560 | rs4702 | $3.5 \times 10^{-9}$ | 1.059 | 0.010 | G/A | 0.446 | FURIN | SCZ, CDG |
| 53   | 16  | 9230816 | rs28455634 | $2.6 \times 10^{-10}$ | 1.065 | 0.010 | G/A | 0.620 | C1orf72 | CDG |
| 54   | 16  | 9926384 | rs7199910 | $1.7 \times 10^{-8}$ | 1.057 | 0.010 | G/T | 0.312 | GRIN2A (PGC2) |  |
| 55   | 16  | 89632725 | rs12932628 | $6.7 \times 10^{-9}$ | 1.058 | 0.010 | T/G | 0.487 | RPL13 |  |
| 56   | 17  | 1835482 | rs4790841 | $3.1 \times 10^{-8}$ | 1.075 | 0.013 | T/C | 0.151 | RTN4RL1 |  |
| 57   | 17  | 38129841 | rs11870683 | $2.8 \times 10^{-8}$ | 1.059 | 0.010 | T/A | 0.650 | ERBB2 (Hou et al. 2016) |  |
| 58   | 17  | 38220432 | rs61554907 | $1.6 \times 10^{-8}$ | 1.091 | 0.015 | T/G | 0.124 | ERBB2 (Hou et al. 2016) |  |
| 59c  | 17  | 42191893 | rs228768 | $2.8 \times 10^{-10}$ | 1.067 | 0.010 | G/T | 0.294 | HDACS (PGC2) |  |
| 60c  | 20  | 43682551 | rs6771285 | $4.2 \times 10^{-11}$ | 1.070 | 0.010 | T/G | 0.687 | STX4 (PGC2) |  |
| 61c  | 20  | 43944323 | rs6032110 | $1.0 \times 10^{-9}$ | 1.059 | 0.009 | A/G | 0.512 | WDFC12 (PGC2) |  |
| 62c  | 20  | 48033127 | rs237460 | $4.3 \times 10^{-9}$ | 1.057 | 0.009 | T/C | 0.412 | KCNB1 | CDG |
| 63   | 20  | 60686581 | rs13044225 | $8.5 \times 10^{-8}$ | 1.056 | 0.010 | G/A | 0.440 | OSPPL2 |  |
| 64   | 22  | 41153879 | rs5758064 | $2.0 \times 10^{-8}$ | 1.054 | 0.009 | T/C | 0.523 | SLC25A17 | MD, SCZ, CDG, MOOD |

\*Previous report refers to previous association of a SNP in the locus with the psychiatric disorder at genome-wide significance. PGC1, ref.1; PGC2, ref.2; Hou et al. 2016, ref.3; Ikeda et al. 2018, ref.4. Green et al. 2013, ref.5; Charney et al. 2017, ref.6. Novel loci are named using the nearest gene to the index SNP. Location overlaps with a genome-wide significant locus for BD I. CHR, chromosome; BP, GRCh37 base-pair position; OR, odds ratio; A1, tested allele; A2, other allele; freq, frequency; CDG, cross-disorder GWAS of the PGC; MD, major depression; SCZ, schizophrenia; MOOD, mood disorders; ASD, autism spectrum disorder; ALC, alcohol use disorder or problematic alcohol use; ADHD, attention deficit/hyperactivity disorder. $P$ values are two-sided and based on an inverse-variance-weighted fixed-effects meta-analysis.
Articles

NATURE GENETICS

Fig. 2 | Phenotypic variance in BD explained by PRSs. Variance explained is presented on the liability scale, assuming a 2% population prevalence of BD. For European ancestries, the results shown are the weighted mean $R^2$ values across all 57 cohorts in the PGC3 meta-analysis, weighted by the effective $n$ per cohort. The numbers of cases and controls are shown from left to right under the bar plot for each study. GWAS $P_T$, the color of the bars represents the $P$-value threshold used to select SNPs from the discovery GWAS; GAIN-AA, Genetic Association Information Network African American cohort; AA-GPC, African American Genomic Psychiatry Cohort.

and smoking initiation, cigarettes per day, problematic alcohol use and drinks per week (Fig. 3). BD was also positively genetically correlated with measures of sleep quality (daytime sleepiness, insomnia and sleep duration; Fig. 3). Among 514 traits measured in the general population of the UK Biobank, there was significant $r_g$ between BD and many psychiatric-relevant traits or symptoms, dissatisfaction with interpersonal relationships, poorer overall health rating and feelings of loneliness or isolation (Supplementary Table 16).

Bivariate gaussian mixture models were applied to the GWAS summary statistics for BD and other complex traits using the MiXeR tool to estimate the number of variants influencing each trait that explain 90% of $h^2_{SNP}$ and their overlap between traits. MiXeR estimated that approximately 8,600 (s.e. = 200) variants influence BD, which is similar to the estimate for schizophrenia (9,700, s.e. = 200) and lower than that for major depression (12,300, s.e. = 600; Supplementary Table 17 and Supplementary Fig. 9). When considering the number of shared loci as a proportion of the total polygenicity of each trait, the vast majority of loci influencing BD were also estimated to influence major depression (97%) and schizophrenia (96%; Supplementary Table 17 and Supplementary Fig. 9). Interestingly, within these shared components, the variants that influenced both BD and schizophrenia had high concordance in direction of effect (80%, s.e. = 2%), while the portion of concordant variants between BD and major depressive disorder was only 69% (s.e. = 1%; Supplementary Table 17).

Genetic and causal relationships between BD and modifiable risk factors. Ten traits associated with BD from clinical and epidemiological studies were investigated in detail for genetic and potentially causal relationships with BD via LDSC, generalized summary statistics–based Mendelian randomization (GSIMR) and bivariate gaussian mixture modeling. BD has been strongly linked with sleep disturbances, alcohol use, smoking, higher educational attainment and mood instability. Most of these traits had modest but significant genetic correlations with BD ($r_g = -0.05$ to 0.35; Fig. 3). Examining the effects of these traits on BD via GSIMR, smoking initiation was associated with BD, corresponding to an OR of 1.49 (95% CI 1.38–1.61) for developing the disorder ($P = 1.74 \times 10^{-22}$, Fig. 3). Testing the effect of BD on the traits, we found that BD was significantly associated with reduced likelihood of being a morning person and increased number of drinks per week ($P < 1.47 \times 10^{-5}$; Fig. 3). Positive bidirectional relationships were identified between BD and longer sleep duration, problematic alcohol use, educational attainment and mood instability (Fig. 3). Notably, the instrumental variables for mood instability were selected from a GWAS conducted in the general population, excluding individuals with psychiatric disorders. For all of the aforementioned BD–trait relationships, the effect size estimates from GSIMR were consistent with those calculated using the inverse-variance-weighted regression method, and there was no evidence of bias from horizontal pleiotropy. Full MR results are in Supplementary Tables 18 and 19. Bivariate gaussian mixture modeling using MiXeR indicated large proportions of variants influencing both BD and all other traits tested, particularly educational attainment, where approximately 98% of variants influencing BD were estimated to also influence educational attainment. While cigarettes per day was a trait of interest, MiXeR could not model these data due to low polygenicity and heritability, and the effect of cigarettes per day on BD was inconsistent between MR methods, suggesting a violation of MR assumptions (Supplementary Tables 18–20).

BD subtypes. We conducted GWAS meta-analyses of BD I (25,060 cases, 449,978 controls) and BD II (6,781 cases, 364,075 controls). The BD I analysis identified 44 genome-wide significant loci, 31 of which overlapped with genome-wide significant loci from the main BD GWAS (Table 1 and Supplementary Table 21). The remaining 13 genome-wide significant loci for BD I all had $P < 4.0 \times 10^{-3}$ in the main BD GWAS. One genome-wide significant locus was identified in the GWAS meta-analysis of BD II and had a $P < 1.1 \times 10^{-4}$ in the main GWAS of BD (Supplementary Table 21). The $h^2_{SNP}$ estimates on the liability scale for BD I and BD II were 20.9% (s.e. = 0.009, $P = 1.0 \times 10^{-11}$) and 11.6% (s.e. = 0.01, $P = 3.9 \times 10^{-15}$), respectively, assuming a 1% population prevalence of each subtype. These heritability values are significantly different from each other ($P = 2.4 \times 10^{-25}$, block jackknife). The genetic correlation between BD I and BD II was 0.85
In a GWAS of 41,917 BD cases, we identified 64 associated genomic loci, 33 of which are novel discoveries. With a 1.5-fold increase in effective sample size compared with the PGC2 BD GWAS, this study more than doubled the number of associated loci, representing an inflection point in the rate of risk variant discovery. We observed consistent replication of known BD loci, including 28/30 loci from the PGC2 GWAS\textsuperscript{44} and several implicated by other BD GWAS\textsuperscript{15–17}, including a study of East Asian cases\textsuperscript{59}. The 33 novel loci discovered here encompass genes of expected biological relevance to BD, such as the ion-channel-encoding genes CACNB2 and KCNB1. Among the 64 BD loci, 17 have previously been implicated in GWAS of schizophrenia\textsuperscript{60}, and 7 in GWAS of major depression\textsuperscript{61}, representing the first overlap of genome-wide significant loci between the mood disorders. For these genome-wide significant loci shared across disorders, 17/17 and 5/7 of the BD index SNPs had the same direction of effect on schizophrenia and major depression, respectively (Supplementary Table 23). More generally, 50/64 and 62/64 BD loci had a consistent direction of effect on major depression and schizophrenia, respectively, considerably greater than chance (\(P<1.47\times 10^{-14}\)), which is significantly different from 1 (\(P=1.6\times 10^{-3}\)). The genetic correlation of BD I with schizophrenia (\(r_g=0.66\), s.e. = 0.02) was higher than that of BD II (\(r_g=0.54\), s.e. = 0.05), whereas major depression was more strongly genetically correlated with BD II (\(r_g=0.66\), s.e. = 0.05) than with BD I (\(r_g=0.34\), s.e. = 0.03; Supplementary Table 22).

**Discussion**

In a GWAS of 41,917 BD cases, we identified 64 associated genomic loci, 33 of which are novel discoveries. With a 1.5-fold increase in effective sample size compared with the PGC2 BD GWAS, this study more than doubled the number of associated loci, representing an inflection point in the rate of risk variant discovery. We observed consistent replication of known BD loci, including 28/30 loci from the PGC2 GWAS\textsuperscript{44} and several implicated by other BD GWAS\textsuperscript{15–17}, including a study of East Asian cases\textsuperscript{59}. The 33 novel loci discovered here encompass genes of expected biological relevance to BD, such as the ion-channel-encoding genes CACNB2 and KCNB1. Among the 64 BD loci, 17 have previously been implicated in GWAS of schizophrenia\textsuperscript{60}, and 7 in GWAS of major depression\textsuperscript{61}, representing the first overlap of genome-wide significant loci between the mood disorders. For these genome-wide significant loci shared across disorders, 17/17 and 5/7 of the BD index SNPs had the same direction of effect on schizophrenia and major depression, respectively (Supplementary Table 23). More generally, 50/64 and 62/64 BD loci had a consistent direction of effect on major depression and schizophrenia, respectively, considerably greater than chance (\(P<1.47\times 10^{-14}\)), which is significantly different from 1 (\(P=1.6\times 10^{-3}\)). The genetic correlation of BD I with schizophrenia (\(r_g=0.66\), s.e. = 0.02) was higher than that of BD II (\(r_g=0.54\), s.e. = 0.05), whereas major depression was more strongly genetically correlated with BD II (\(r_g=0.66\), s.e. = 0.05) than with BD I (\(r_g=0.34\), s.e. = 0.03; Supplementary Table 22).
variant in **BTN2A1**, a brain-expressed gene encoding a plasma membrane protein.

The genetic correlation of BD with other psychiatric disorders was consistent with previous reports. Our results also corroborate previous genetic and clinical evidence of associations between BD and sleep disturbances, problematic alcohol use and smoking. While the genome-wide genetic correlations with these traits were modest \((r = -0.05\) to \(0.35\)), MiXiR estimated that, for all traits, more than 55% of trait-influencing variants also influence BD (Fig. 3). Taken together, these results point to shared biology as one possible explanation for the high prevalence of substance use in BD. However, excluding genetic variants associated with both traits, MR analyses suggested that smoking is also a putatively 'causal' risk factor for BD, while BD has no effect on smoking, consistent with a previous report. (We use the word 'causal' with caution here as we consider MR an exploratory analysis to identify potentially modifiable risk factors that warrant more detailed investigations to understand their complex relationship with BD.) In contrast, MR indicated that BD had bidirectional 'causal' relationships with problematic alcohol use, longer sleep duration and mood instability. Insights into the relationship of such behavioral correlates with BD may have future impact on clinical decision-making in the prophylaxis or management of the disorder. Higher educational attainment has previously been associated with BD in epidemiological studies, while lower educational attainment has been associated with schizophrenia and major depression. Here, educational attainment had a significant positive effect on risk of BD and vice versa. Interestingly, MiXiR estimated that almost all variants that influence BD also influence educational attainment. The substantial genetic overlap observed between BD and the other phenotypes suggests that many variants likely influence multiple phenotypes, which may be differentiated by phenotype-specific effect size distributions among the shared influencing variants.

The integration of eQTL data with our GWAS results yielded 15 high-confidence genes for which there was converging evidence that their association with BD is mediated via gene expression. Among these were **HTR6**, encoding a serotonin receptor targeted by antipsychotics and antidepressants, and **MCVR** (melanin-concentrating hormone receptor 1), encoding a target of the antipsychotic haloperidol. We note that, for both of these genes, their top eQTLs have opposite directions of effect on gene expression in the brain and blood, possibly playing a role in the tissue-specific gene regulation influencing BD. BD was associated with decreased expression of **FURIN**, a gene with a neurodevelopmental role that has already been the subject of functional genomics experiments in neuronal cells following its association with schizophrenia in GWAS. The top association in our GWAS was in the **TRANK1** locus on chromosome 3, which has previously been implicated in BD. Although BD-associated SNPs in this locus are known to regulate **TRANK1** expression, our eQTL analyses support a stronger but correlated regulation of **DCLK3**, located 87 kilobases (kb) upstream of **TRANK1** (refs. \(1,27\)). Both **FURIN** and **DCLK3** also encode druggable proteins (although they are not targets for any current psychiatric medications)\(7,28\). These eQTL results provide promising BD candidate genes for functional follow-up experiments. While several of these are in genome-wide significant loci, many are not the closest gene to the index SNP, highlighting the value of probing underlying molecular mechanisms to prioritize the most likely causal genes in the loci.

GWAS signals were enriched in the gene targets of existing BD pharmacological agents, such as antipsychotics, mood stabilizers and antiepileptics. However, enrichment was also found in the targets of calcium channel blockers used to treat hypertension and GABA-receptor-targeting anesthetics (Supplementary Table 8). Calcium channel antagonists have long been investigated for the treatment of BD, without becoming an established therapeutic approach, and there is evidence that some antiepileptics have calcium-channel-inhibiting effects. These results underscore the opportunity for repurposing some classes of drugs, particularly calcium channel antagonists, as potential BD treatments.

BD associations were enriched in gene sets involving neuronal parts and synaptic signaling. Neuronal and synaptic pathways have been described in cross-disorder GWAS of multiple psychiatric disorders including BD\(12-24\). Disregulation of such pathways has also been suggested by previous functional and animal studies. Analysis of single-cell gene expression data revealed enrichment in genes with high specificity of gene expression in neurons (both excitatory and inhibitory) of many brain regions, in particular the cortex and hippocampus. These findings are similar to those reported in GWAS data of schizophrenia and major depressive disorder.

PRSs for BD explained on average 4.57% of phenotypic variance (liability scale) across European cohorts, although this varied in different waves of the BD GWAS, ranging from 6.6% in the PGC1 cohorts to 2.9% in the American biobanks (Supplementary Table 12). These results are in line with the higher PRSs of BD per wave, which ranged from 24.6% (s.e. = 0.01) in PGC1 to 11.9% (s.e. = 0.01) in external studies (Supplementary Table 3). Some variability in the BD PRS estimates may arise from the inclusion of cases from population biobanks, which may have more heterogeneous clinical presentations or less severe illness than patients with BD ascertained via inpatient or outpatient psychiatric clinics. Across the waves of clinically ascertained samples within the PGC, the BD PRSs are also higher, likely reflecting clinical and genetic heterogeneity in the type of BD cases ascertained; the PGC1 cohorts consisted mostly of BD I cases, known to be the most heritable of the BD subtypes\(2,24\), while later waves included more individuals with BD II. Overall, the BD PRS calculated from the meta-analysis summary statistics was 18% on the liability scale, a decrease of ~2% compared with the PGC2 GWAS, which may be due to the addition of cohorts with lower PRSs and heterogeneity between cohorts. However, despite differences in the BD PRSs and the higher PRSs per wave, the genetic correlation of BD between all waves was high (weighted mean \(r_g = 0.94\), s.e. = 0.03), supporting our rationale for combining cases with different BD subtypes or ascertainment to increase power for discovery of risk variants. In Europeans, individuals in the top 10% of PRSs had an OR of 3.5 for BD, compared with individuals with average PRSs (middle decile), which translates into a modest lifetime risk of the disorder (7% based on PRS alone). While PRSs are invaluable tools in research settings, the current BD PRS lacks sufficient power to separate individuals into clinically meaningful risk categories, and therefore have no clinical utility at present\(2,28\). PRSs from this European BD meta-analysis yield higher \(R^2\) values in diverse ancestry samples than PRSs based on currently available BD GWAS within the same ancestry\(99\). However, performance still greatly lags behind that in Europeans, with ~2% variance explained in East Asian samples and substantially less in admixed African American samples, likely due to differences in allele frequencies and LD structures, consistent with previous studies\(99,100\). There is a pressing need for more and larger studies in other ancestry groups to ensure that any future clinical utility is broadly applicable. Exploiting the differences in LD structure between diverse ancestry samples will also assist in the fine-mapping of risk loci for BD.

Our analyses confirmed that BD is a highly polygenic disorder, with an estimated 8,600 variants explaining 90% of its BD genetic load. Hence, many more SNPs than those identified here are expected to account for the common variant architecture underlying BD. This GWAS marks an inflection point in risk variant discovery, and we expect that, from this point forward, the addition of more samples will lead to a dramatic increase in genetic findings. Nevertheless, fewer genome-wide significant loci have been identified in BD than in schizophrenia GWAS of comparable sample size. This may be due to the clinical and genetic heterogeneity that exists in BD.
Our GWAS of subtypes BD I and BD II identified additional associated loci. Consistent with previous findings, our analysis showed that the two subtypes were highly but imperfectly genetically correlated ($r = 0.85$), and that BD I is more genetically correlated with schizophrenia, while BD II has stronger genetic correlation with major depression. The subtypes are sufficiently similar to justify joint analysis as BD, but are not identical in their genetic composition, and as such contribute to the genetic heterogeneity of BD. We identified 13 loci passing genome-wide significance for BD I, and one for BD II, which did not reach significance in the main BD GWAS, further illustrating the partially differing genetic composition of the two subtypes. Understanding the shared and distinct genetic components of BD subtypes and symptoms requires detailed phenotyping efforts in large cohorts and is an important area for future psychiatric genetics research.

In summary, these new data advance our understanding of the biological etiology of BD and prioritize a set of candidate genes for functional follow-up experiments. Several lines of evidence converge on the involvement of calcium channel signaling, providing a promising avenue for future therapeutic development.

Online content
Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-021-00857-4.

Received: 2 September 2020; Accepted: 25 March 2021; Published online: 17 May 2021

References
1. GBD 2016 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet 390, 1211–1259 (2017).
2. Plans, L. et al. Association between completed suicide and bipolar disorder: a systematic review of the literature. J. Affect. Disord. 242, 111–122 (2019).
3. American Psychiatric Association Diagnostic and Statistical Manual of Mental Disorders 5th edn (American Psychiatric Association Publishing, 2013).
4. Merikangas, K. R. et al. Lifetime and 12-month prevalence of bipolar spectrum disorder in the national comorbidity survey replication. Arch. Gen. Psychiatry 64, 543–552 (2007).
5. Merikangas, K. R. et al. Prevalence and correlates of bipolar spectrum disorder in the world mental health survey initiative. Arch. Gen. Psychiatry 68, 241–251 (2011).
6. Craddock, N. & Sklar, P. Genetics of bipolar disorder. Lancet 381, 1654–1662 (2013).
7. Song, J. et al. Bipolar disorder and its relation to major psychiatric disorders: a family-based study in the Swedish population. Bipolar Disord. 17, 184–193 (2015).
8. Bienvenu, O. J., Davydow, D. S. & Kendler, K. S. Psychiatric ‘diseases’ versus behavioral disorders and degree of genetic influence. Psychol. Med. 41, 33–40 (2011).
9. Psychiatric GWAS Consortium Bipolar Disorder Working Group. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. Nat. Genet. 43, 977–983 (2011).
10. Baum, A. E. et al. A genome-wide association study implicates diacylglycerol kinase eta (DGKH) and several other genes in the etiology of bipolar disorder. Mol. Psychiatry 13, 197–207 (2008).
11. Charney, A. W. et al. Evidence for genetic heterogeneity between clinical subtypes of bipolar disorder. Transl. Psychiatry 7, 6993 (2017).
12. Chen, D. T. et al. Genome-wide association study meta-analysis of European and Asian-ancestry samples identifies three novel loci associated with bipolar disorder. Mol. Psychiatry 18, 195–205 (2013).
13. Cicchón, S. et al. Genome-wide association study identifies genetic variation in neurocan as a susceptibility factor for bipolar disorder. Am. J. Hum. Genet. 88, 372–381 (2011).
14. Ferreira, M. A. R. et al. Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. Nat. Genet. 40, 1056–1058 (2008).
15. Green, E. K. et al. Association at SYNE1 in both bipolar disorder and recurrent major depression. Mol. Psychiatry 18, 614–617 (2013).
16. Green, E. K. et al. Replication of bipolar disorder susceptibility alleles and identification of two novel genome-wide significant associations in a new bipolar disorder case–control sample. Mol. Psychiatry 18, 1302–1307 (2013).
17. Hou, L. et al. Genome-wide association study of 40,000 individuals identifies two novel loci associated with bipolar disorder. Hum. Mol. Genet. 25, 3383–3394 (2016).
18. Michielesen, T. W. et al. Genome-wide association study reveals two new risk loci for bipolar disorder. Nat. Commun. 5, 3339 (2014).
19. Schulze, T. G. et al. Two variants in Ankyrin 3 (ANK3) are independent genetic risk factors for bipolar disorder. Mol. Psychiatry 14, 487–491 (2009).
20. Scott, L. J. et al. Genome-wide association and meta-analysis of bipolar disorder in individuals of European ancestry. Proc. Natl Acad. Sci. USA 106, 7591–7596 (2009).
21. Sklar, P. et al. Whole-genome association study of bipolar disorder. Mol. Psychiatry 13, 558–569 (2008).
22. Smith, E. N. et al. Genome-wide association study of bipolar disorder in European American and African American individuals. Mol. Psychiatry 14, 56–763 (2009).
23. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447, 661–678 (2007).
24. Stahl, E. A. et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. Nat. Genet. 51, 793–803 (2019).
25. Lee, S.-H., Zabolotny, J. M., Huang, H., Lee, H. & Kim, Y.-R. Insulin in the nervous system and the mind: functions in metabolism, memory, and mood. Mol. Metab. 5, 589–601 (2016).
26. McIntyre, R. S. et al. A randomized, double-blind, controlled trial evaluating the effect of intranasal insulin on neurocognitive function in euthymic patients with bipolar disorder. Bipolar Disord. 14, 697–706 (2012).
27. Nurnberger, J. I. Jr et al. Identification of pathways for bipolar disorder: a meta-analysis. JAMA Psychiatry 71, 657–664 (2014).
28. Gordonove, F. J. A. & McMahon, F. J. The genetics of bipolar disorder. Mol. Psychiatry 25, 544–559 (2020).
29. Zhang, C., Xiao, X., Li, T. & Li, M. Translational genomics and beyond in bipolar disorder. Mol. Psychiatry 26, 186–202 (2021).
30. Pedersen, C. B. et al. The iPSYCH2012 case–cohort sample: new directions for unravelling genetic and environmental architectures of severe mental disorders. Mol. Psychiatry 1, 6–14 (2018).
31. Gudbjartsson, D. F. et al. Large-scale whole-genome sequencing of the Icelandic population. Nat. Genet. 47, 435–444 (2015).
32. Leitsalu, L. et al. Cohort Profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. Int. J. Epidemiol. 44, 1137–1147 (2015).
33. Krokstad, S. et al. Cohort Profile: the HUNT Study, Norway. Int. J. Epidemiol. 42, 968–977 (2013).
34. Sudlow, C. et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 12, e1001779 (2015).
35. Bulik-Sullivan, B. K. et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat. Genet. 47, 291–295 (2015).
36. Loh, P.-R., Kichaev, G., Gazal, S., Schoech, A. P. & Price, A. L. Mixed-model association for biobank-scale datasets. Nat. Genet. 50, 906–908 (2018).
37. de Leeuw, C. A., Mooij, J. M., Heskens, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput. Biol. 11, e1004219 (2015).
38. Bryois, J. et al. Genetic identification of cell types underlying brain complex traits yields insights into the etiology of Parkinson’s disease. Nat. Genet. 52, 482–493 (2020).
39. Wagner, A. H. et al. DGlbd 2.0: mining clinically relevant drug–gene interactions. Nucleic Acids Res. 44, D1036–D1044 (2016).
40. Roth, B. L., Lopez, E., Patel, S. & Kroeeze, W. K. The multiplicity of serotonin receptors: uselessly diverse molecules or an embarrassment of riches? Neuron Sci. 6, 252–262 (2000).
41. Gaspar, H. A. & Brenn, G. Drug enrichment and discovery from schizophrenia genome-wide association results: an analysis and visualisation approach. Sci. Rep. 7, 12460 (2017).
42. Gusev, A. et al. Integrative approaches for large-scale transcriptome-wide association studies. Nat. Genet. 48, 245–252 (2016).
43. Gandal, M. J. et al. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. Science 362, eaat8127 (2018).
44. Mancuso, N. et al. Probabilistic fine-mapping of transcriptome-wide association studies. Nat. Genet. 51, 675–682 (2019).
45. Zhu, Z. et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat. Genet. 48, 481–487 (2016).
46. Wu, Y. et al. Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. Nat. Commun. 9, 918 (2018).
47. Vösa, U. et al. Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. Preprint at bioRxiv https://doi.org/10.1101/447367 (2018).

48. Zheng, J. et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. Bioinformatics 33, 272–279 (2017).

49. Frei, O. et al. Bivariate causal mixture model quantifies polygenic overlap between complex traits beyond genetic correlation. Nat. Commun. 10, 2417 (2019).

50. Holland, D. et al. Beyond SNP heritability: polygenicity and discoverability of phenotypes estimated with an univariate Gaussian mixture model. PLoS Genet. 16, e1008612 (2020).

51. Zhu, Z. et al. Causal associations between risk factors and common diseases inferred from GWAS summary data. Nat. Commun. 9, 224 (2018).

52. Steardo, L. Jr et al. Sleep disturbance in bipolar disorder: neuroglia and neurobiological dysfunction. J. Affect. Disord. 206, 331–349 (2016).

53. Hunt, G. E., Malhi, G. S., Cleary, M., Lai, H. M. X. & Satharthan, T. Prevalence of comorbid bipolar and substance use disorders in clinical settings from 1990–2019: systematic review and meta-analysis. J. Affect. Disord. 207, 283–285 (2015).

54. Vreeker, A. et al. High educational performance is a distinctive feature of individuals with bipolar disorder I or II and control participants. JAMA Psychiatry 75, 1469–1482 (2018).

55. MacCabe, J. H. et al. Excellent school performance at age 16 and risk of adult bipolar disorder: national cohort study. Br. J. Psychiatry 196, 89–115 (2010).

56. Broome, M. R., Saunders, K. E. A., Harrison, P. J. & Marwaha, S. Mood instability: significance, definition and measurement. Br. J. Psychiatry 207, 283–285 (2015).

57. Lewis, K. J. et al. Comparison of genetic liability for sleep traits among psychiatric disorders. Neuropsychopharmacol. 42, 366–379 (2017).

58. Mizuno, A. & Okada, Y. Biological characterization of expression quantitative trait loci (eQTLs) showing tissue-specific opposite directional effects. Eur. J. Hum. Genet. 27, 1745–1756 (2019).

59. Jiang, X. et al. Sodium valproate rescues expression of TRANK1 in iPSC-derived neural cells that carry a genetic variant associated with serious mental illness. Mol. Psychiatry 24, 613–624 (2019).

60. Finan, C. et al. The druggable genome and support for target identification and validation in drug development. Sci. Transl. Med. 9, eaag1166 (2017).

61. von Wegerer, J., Hesslinger, B., Berger, M. & Walden, J. A calcium antagonistic effect of the new antiepileptic drug lamotrigine. Eur. Neuropsychopharmacol. 10, 77–81 (1997).

62. Cipriani, A. et al. A systematic review of calcium channel antagonists in bipolar disorder and some considerations for their future development. Mol. Psychiatry 21, 1324–1332 (2016).

63. Harrison, P. J., Tunbridge, E. M., Dolphin, A. C. & Hall, J. Voltage-gated calcium channel blockers for psychiatric disorders: genomic reappraisal. Br. J. Psychiatry 216, 250–253 (2020).

64. Network and Pathway Analysis Subgroup of Psychiatric Genomics Consortium. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. Nat. Neurosci. 18, 199–209 (2015).

65. Forster, A. J. et al. Identification of shared risk loci and pathways for bipolar disorder and schizophrenia. PLoS ONE 12, e0173595 (2017).

66. Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium. Genomic dissection of bipolar disorder and schizophrenia, including 28 subphenotypes. Cell 173, 1705–1715 (2018).

67. Lee, Y., Zhang, Y., Kim, S. & Han, K. Excitatory and inhibitory synaptic dysfunction in mania: an emerging hypothesis from animal model studies. Exp. Mol. Med. 50, 1–11 (2018).

68. Skene, N. G. et al. Genetic identification of brain cell types underlying schizophrenia. Nat. Genet. 50, 825–833 (2018).

69. Lewis, C. M. & Vassos, E. Polygenic risk scores: from research tools to clinical instruments. Genome Med. 12, 44 (2020).

70. Torkamani, A., Winearling, N. E. & Topol, E. J. The personal and clinical utility of polygenic risk scores. Nat. Rev. Genet. 19, 581–590 (2018).

71. Duncan, L. et al. Analysis of polygenic risk score usage and performance in diverse human populations. Nat. Commun. 10, 3328 (2019).

72. Martin, A. et al. Clinical use of current polygenic risk scores may exacerbate health disparities. Nat. Genet. 51, 584–591 (2019).

73. Coleman, J. R. I. et al. The genetics of the mood disorder spectrum: genome-wide association analyses of more than 185,000 cases and 439,000 controls. Biol. Psychiatry 88, 169–184 (2020).

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. © The Author(s), under exclusive licence to Springer Nature America, Inc. 2021
Adebayo Anjorin44, Verneri Antilla45, Anastasia Antoniou46, Swapnil Awasthi48, Ji Hyun Baek47, Marie Bækvad-Hansen14,48, Nicholas Bass40, Michael Bauer49, Eva C. Beins3, Sarah E. Bergen20, Armin Birner50, Carsten Bøcker Pedersen14,38,39, Erlend Bøen51, Marco P. Boks52, Rosa Bosch53,54,55,56, Murielle Brun57, Ben M. Brumpton19, Nathalie Brunhorst-Kanaan57, Monika Budde36, Jonas Bybjerg-Grauholm14,48, William Byerley58, Murray Cairns59, Miquel Casas53,54,55,56, Pablo Cervantes60, Toni-Kim Clarke61, Cristiana Cruceanu60,62, Alfredo Cuellar-Barboza43,64, Julie Cunningham65, David Curtis66,67, Piotr M. Czerski68, Anders M. Dale69, Nina Dalkner50, Friederike S. David8, Franziska Degenhardt3,70, Srdjan Djurovic71,72, Amanda L. Dobbs1,2, Athanassios Douzenis46, Torbjørn Elvsåshagen18,73,74, Valentina escott-Price75, I. Nicol Ferrier76, Alessia Fiorentino40, Tatiana M. Foroud77, Liz Forty75, Josef Frank78, Oleksandr Frei16,18, Nelson B. Freimer23,79, Louise Frisen60, Katrin Gade36,81, Julie Garnham82, Joel Gerlenter93,84,85, Marianne Giørtz Pedersen14,38,39, Ian R. Gizer86, Scott D. Gordon87, Katherine Gordon-Smith88, Tiffany A. Greenwood89, Jakob Grove12,13,14,90, José Guzman-Parral91, Kyosoo Ha92, Magnus Haraldsson93, Martin Hautzinger94, Urs Heilbronner36, Dennis Hellgren20, Stefan Herms3,95,96, Per Hoffmann95,96, Peter A. Holmans79, Laura Huckins1,2, Stéphane Jamain97,98, Jessica S. Johnson1,2, Janos L. Kalman36,37,99, Yoichiro Kamatani100,101, James L. Kennedy102,103,104,105, Sarah Kittel-Schneider57,106, James A. Knowles107,108, Manolis Kogevinas109, Maria Koromina110, Thorsten M. Kranz57, Henry R. Kranzler111,112, Michiaki Kubo113, Ralph Kupca114,115,116, Steven A. Kushner117, Catharina Lavebatt118,119, Jacob Lawrence120, Markus Leber21, Heon-Jeong Lee122, Phil H. Lee123, Shawn E. Levy124, Catrin Lewis75, Calwing Liao125,126, Susanne Lucas62, Martin Lundberg118,119, Donald J. MacIntyre127, Sigurdur H. Magnusson32, Wolfgang Maier128, Adam Maihofer89, Dolores Malaspina1,2, Eirini Maratou129, Lina Martinsson80, Manuel Mattheisen12,13,14,106,130, Steven A. McCarroll25,26, Nathaniel W. McGregor131, Peter McGuffin9, James D. McKay132, Helena Medeiros108, Sarah E. Medland87, Vincent Millischer118,119, Grant W. Montgomery11, Jennifer L. Moran25,133, Derek W. Morris134, Thomas W. Mühlisen4,95, Niamh O’Brien40, Claire O’Donovan82, Loes M. Olde Loohuis23,79, Lilijana Oruc135, Sergi Papiol36,37, Antonio F. Pardiñas75, Amy Perry88, Andrea Pfennig49, Evgenia Porichi46, James B. Potash136, Digby Quested137,138, Twofique Raj1,29,30,31, Mark H. Rapaport139, J. Raymond DePaulo136, Eline J. Regeer140, John P. Rice141, Fabio Rivas91, Margarita Rivera142,143, Julian Roth106, Panos Roussos1,2,29, Douglas M. Ruderfer144, Cristina Sánchez-Mora13,53,54,56,145, Eva C. Schulte36,37, Fanny Senner36,37, Sally Sharp40, Paul D. Shilling89, Engilbert Sigurdsson93,146, Lea Sirignano78, Claire Slaney82, Olav B. Smeland6,7, Daniel J. Smith147, Janet L. Sobell148, Christine Søholm Hansen14,48, Maria Soler Artigas53,54,56,145, Anne T. Spijker149, Dan J. Stein150, John S. Strauss102, Beata Świątkowska151, Chikashi Terao101, Thorgeir E. Thorgeirsson42, Claudio Tomai152,153,154, Paul Tooney49, Evangelia-Eirini Tsermpini110, Marquis P. Vawter155, Helmut Vedder156, James T. R. Walters75, Stephanie H. Witt78, Simon Xi157, Wei Xu158, Jessica Mei Kay Yang75, Allan H. Young159,160, Hannah Young1, Peter P. Zandi136, Hang Zhou43,84, Lea Zillich78, HUNT All-In Psychiatry, Rolf Adolfsson161, Ingrid Agartz51,130,162, Martin Alda82,163, Lars Alfredsson164, Gulja Babadjanova165, Lena Backlund118,119, Bernhard T. Baune166,167,168, Frank Bellivier169,170, Susanne Bengesser50, Wade H. Berrettini171, Douglas H. R. Blackwood81, Michael Boehnke172, Anders D. Berglund14,173,174, Gerome Breen9,10, Vaughan J. Carr175, Stanley Catts176, Aiden Corvin177, Nicholas Craddock75, Udo Dannlowski166, Dimitris Dikeos178, Tònu Esko26,27,179,180, Bruno Etain169,170, Panagiotis Ferentinos46,9, Mark Frye54, Janice M. Fullerton152,153, Micha Gawlik106, Elliot S. Gershon42,181,
Fernando S. Goes, Melissa J. Green, Maria Grigorieu-Serbanescu, Joanna Hauser, Frans Henskens, Jan Hillert, Kyung Sue Hong, David M. Hougaard, Christina M. Hultman, Kristian Hveem, Nakao Iwata, Assen V. Jablensky, Ian Jones, Lisa A. Jones, René S. Kahn, John R. Kelsoe, George Kirou, Mikael Landén, Marion Leboyer, Cathryn M. Lewis, Qingqin S. Li, Jolanta Lissowska, Christine Lochner, Carmel Loughland, Nicholas G. Martin, Carol A. Mathews, Fermin Mayoral, Susan L. McElroy, Andrew M. McIntosh, Francis J. McMahon, Ingrid Melle, Patricia Michie, Lili Milani, Philip B. Mitchell, Gunnar Morken, Ole Mors, Preben Bo Mortensen, Bryan Mowry, Bertram Müller-Myhso, Richard M. Myers, Benjamin M. Neale, Caroline M. Nievergelt, Merete Nordentoft, Markus M. Nöthen, Michael C. O’Donovan, Ketil J. Oedegaard, Tomas Olsson, Michael J. Owen, Sara A. Paciga, Chris Pantelis, Carlos Pato, George P. Patrinos, Roy H. Perlis, Danielle Posthuma, Josep Antoni Ramos-Quiroga, Andreas Reif, Eva Z. Reininghaus, Marta Ribasés, Marcella Rietschel, Stephan Ripke, Guy A. Rouleau, Takeo Saito, Ulrich Schall, Martin Schalling, Peter R. Schofield, Thomas G. Schulze, Laura J. Scott, Rodney J. Scott, Alessandro Serretti, Cynthia Shannon Weickert, Jordan W. Smoller, Kari Stefansson, Eystein Stordal, Fabian Streit, Patrick F. Sullivan, Gustavo Turecki, Arne E. Vaaler, Eduard Vieta, John B. Vincent, Irwin D. Waldman, Thomas W. Weickert, Thomas Werge, Naomi R. Wray, John-Arker Zwart, Joanna M. Biernacka, Preben Bo Mortensen, Bryan Mowry, Bertram Müller-Myhso, Richard M. Myers, Thomas Werge, Naomi R. Wray, John-Arker Zwart, Joanna M. Biernacka, John I. Nurnberger, Sven Cichon, Howard J. Edenberg, Eli A. Stahl, Andrew McQuillain, Arianna Di Florio, Roel A. Ophoff, and Ole A. Andreassen.
Prentice et al. 2016 | Nature Genetics | Vol 48 | Issue 5 | Pages 576-582 | DOI: 10.1038/ng.3600

A list of members and their affiliations appears in the Supplementary Information.

HUNT All-In Psychiatry

Bendik S. Winsvold19,33, Eystein Stordal220,221, Gunnar Morken21,197, John-Anker Zwart18,19,33, Ole Kristian Drange21,22 and Sigrid Børte17,18,19

A list of authors and their affiliations appears in the Supplementary Information.
Methods

Sample description. The meta-analysis sample comprises 57 cohorts collected in Europe, North America and Australia, totaling 41,917 BD cases and 371,549 controls of European descent (Supplementary Table 1). The total effective sample size, an equal number of cases and controls in each cohort ($4(n_{cases} \times n_{controls})/(n_{cases} + n_{controls})$, is 101,962. For 52 cohorts, individual-level genotype and phenotype data were shared with the PGC. Cohorts have been added to the PGC in five waves (PGC1 (ref. 97), PGC2 (ref. 98), PGC PsychChip, PGC3 and External studies); all cohorts from previous PGC BD GWASs were included. The source and inclusion/exclusion criteria for cases and controls for each cohort are described in the Supplementary Note. Cases were required to meet international consensus criteria (DSM-IV, ICD-9 or ICD-10) for a lifetime diagnosis of BD, established using structured diagnostic instruments from assessments by trained interviewers, clinician-administered checklists or medical record review. In most cohorts, controls were screened for the absence of lifetime psychiatric disorders and randomly selected from the population. For five cohorts (iPSYCH30, deCODE and one of each pair of related individuals (pi_hat using EIGENSTRAT v6.1.4 (ref. 95). On the basis of visual inspection of plots of components (PCs) were generated using genotyped SNPs in each cohort separately and Illumina). Subsequently, standardized quality control, imputation and using standard genotype calling software from commercial sources (Affymetrix n effective

Overlap of loci with other psychiatric disorders. Genome-wide significant loci for BD were assessed for overlap with genome-wide significant loci for other psychiatric disorders, using the largest available GWAS results for major depression (schizophrenia, autism spectrum disorders, post-traumatic stress disorder, lifetime anxiety disorder, Tourette syndrome, anorexia nervosa, alcohol use disorder or problematic alcohol use, autism spectrum disorder, mood disorders and the cross-disorder GWAS of the PGC8). The boundaries of the genome-wide significant loci were calculated in the original publications. Overlap of loci was calculated using bedtools v2.29.2 (ref. 92).

Enrichment analyses. P values quantifying the degree of association of genes and gene sets with BD were calculated using MAGMA v1.08 (ref. 97), implemented in FUMA v1.3.6a (refs. 98, 99). Gene-based tests were performed for 19,576 genes (Bonferroni-corrected P-value threshold = 2.55 × 10^{-4}). A total of 11,858 curated gene sets, including at least 10 genes per set (minimum overlap with BD) were imputed to the HRC reference panel (Bonferroni-corrected P-value threshold = 4.22 × 10^{-4}). Competitive gene-set tests were conducted correcting for gene size, variant density and LD within and between genes. Tissue-set enrichment analyses were also performed using MAGMA implemented in FUMA, to test for enrichment of association signal in genes expressed in 54 tissue types from GTEx V8 (Bonferroni-corrected P-value threshold = 9.26 × 10^{-4}). For single-cell enrichment analyses, publicly available single-cell RNA-sequencing data were compiled from five studies of the adult human and mouse brain89. The mean expression for each gene in each cell type was calculated from the single-cell data (not of the single-cell dataset100, we used the mean expression at level 4 (39 cell types from 19 regions for the mouse nervous system). For the Saunders dataset101, we computed the mean expression of the different classes in each of the 9 different brain regions sampled (88 cell types in total). We filtered out any genes with nonunique names, genes not expressed in any cell types, non-protein-coding genes and, for mouse datasets, genes that were no expert-curated gene and no ortholog between mouse and human (Mouse Genome Informatics, The Jackson Laboratory, version 11/22/2016, http://www. informatics.jax.org/downloads/reports/index.html?homology), resulting in 16,472 genes. Expression was then scaled to a total of 1 million unique molecular identifiers (or transcripts per million) for each cell type/tissue. Using a previously described method90, we calculated a metric of gene expression specificity by dividing the expression of each gene in each cell type by the total expression of that gene in all cell types, leading to values ranging from 0 to 1 for each gene (0 meaning that the gene is not expressed in that cell type and 1 meaning that all of the expression of the gene is in that cell type). We then selected the top 10% most specific genes for each cell type/tissue for enrichment analysis. MAGMA v1.08 (ref. 97) was used to test gene-set enrichment using GWAS summary statistics, covarying for gene size, gene density, mean sample size for tested SNPs per gene, the inverse of the minor allele counts per gene and the log of these metrics. We excluded any SNPs with INFO score < 0.6, with MAF < 1% or with estimated odds ratio > 25 or smaller than 1/25, as well as SNPs located in the MHC region (chr6:25–34 Mb). We set a window of 35 kb upstream to 10 kb downstream of the gene coordinates to compute gene-level association statistics and used the European reference panel from phase 3 of the 1000 Genomes Project as the reference population111. We then used MAGMA to test whether the 10% most specific genes (with an expression of at least 1 transcript per million or 1 unique mouse identifier per million) for each cell type/tissue were associated with BD. The P-value threshold for significance was P < 0.01, representing a 5% false discovery rate across datasets.

Further gene-set analyses were performed restricted to genes targeted by drugs, assessing individual drugs and grouping drugs with similar actions. This approach has been described previously92. Gene-level and gene-set analyses were performed in MAGMA v1.08 (ref. 97). Gene boundaries were defined using build 37 reference data from the National Center for Biotechnology Information, available on the MAGMA website (https://ctg.cncr.nl/software/magma), extended 35 kb upstream and 10 kb downstream to include regulatory regions outside the transcript region. Gene-level association statistics were defined as the aggregate of the mean and the lowest level P value within the gene boundaries to calculate the gene-wide P value. Gene sets were defined comprising the targets of each drug in the Drug–Gene Interaction database DGI v2.2 (ref. 37) and in the Psychoactive Drug Screening Database Ki DB v8, both downloaded in June 201654. Analyses were performed using competitive gene-set analyses in MAGMA. Results from the drug-set analysis were then grouped according to the Anatomical Therapeutic Chemical class of the drug93. Only drug classes with at least ten valid drug gene sets within them were analyzed. Drug-class analysis was performed using enrichment curves. All drug gene sets were ranked by their association in the drug-set analysis, and then for a given drug class an enrichment curve was drawn scoring a ‘hit’ if the drug gene set was within the class, or a ‘miss’ if it was outside the class. The area under the curve was calculated, and a genome-wide significant locus for this was calculated using the Bonferroni–Mann–Whitney test comparing drug gene sets within the class to drug gene sets outside the class104. Multiple testing was controlled using a Bonferroni-corrected significance threshold of P < 5.60 × 10^{-4} for drug-set analysis and P < 7.93 × 10^{-4} for drug-class analysis, accounting for 893 drug sets and 63 drug classes tested.
Articles NATURE GENETICS

ε

N

+ = 0

H

NATURE GENETICS was predicted from the imputed C4A C4 AL), rs13195402 genotype (top lead SNP in the MHC) and PCs as per the GW AS. C4 2% and 1%. LDSC bivariate genetic correlations attributable to genome-wide SNPs P least one significant external cohorts by the collaborating research teams using comparable procedures. Covarying for PCs as per the GW AS in each cohort. PRSs were tested in the results of TW AS and SMR indicate an association between BD and gene expression weights from PsychENCODE data (1,321 brain samples)43, available A TW AS was conducted using the precomputed gene expression, rs13195402 genotype (top lead SNP in the MHC) and PCs as per the TW AS fine-mapping of the region was conducted using FOCUS (fine-mapping with two key exceptions: we used the GRCh37 PLINK recombination map, and we with two key exceptions: we used the GRCh37 PLINK recombination map, and we set the output to include genotype probability (that is, GP field in VCF) for correct downstream probabilistic estimation of C4A and C4B joint dosages. The output consisted of dosage estimates for each of the common C4 structural haplotypes for each individual. The highest frequency common structural forms of the C4A/C4B loci (BS, AL, AL–BS, AL–BL, and AL–AL) could be inferred with reasonably high accuracy (generally 0.70 < P < 1.00). The imputed C4 alleles were tested for association with BD in a joint logistic regression that included terms for dosages of the five most common C4 structural haplotypes (AL–BS, AL–BL, and AL–AL). rs13195402 genotype (top lead SNP in the MHC) and PCs as per the GWAS. The genotypically regulated expression of C4A was predicted from the imputed C4 alleles using a model previously described.43 Predicted C4A expression was tested for association with BD in a joint logistic regression that included predicted C4A expression, rs13195402 genotype (top lead SNP in the MHC) and PCs as per the GWAS.

Polygenic risk scoring. PRSs from our GWAS meta-analysis were tested for association with BD in individual cohorts, using a discovery GWAS where the target cohort was left out of the meta-analysis. Briefly, the GWAS results from each discovery GWAS were pruned for LD using the P-value-informed clipping method in PLINK v1.90 (ref.; P < 0.1 within a 300 kb window) based on the LD structure of the HRC reference panel.44 Subsets of SNPs were selected from the results below nine increasingly liberal P-value thresholds (GWAS P; 5 × 10–4, 1 × 10–4, 1 × 10–5, 1 × 10–5, 0.01, 0.05, 0.1, 0.2, 0.5, 1). Sets of alleles, weighted by their log odds ratios from the discovery GWAS, were summed into PRSs for each individual in the target datasets, using PLINK v1.90 implemented via RCOPILE14. PRSs were tested for association with BD in the discovery GWAS with parameter estimation using logistic regression, covarying for PCs as per the GWAS in each cohort. PRSs were tested in the external cohorts by the collaborating research teams using comparable procedures. The variance explained by the PRSs (R2) was converted to the liability scale to account for the proportion of cases in each target dataset, using a BD population prevalence of 2% and 1%.15 The weighted average R2 values were calculated using the effective n for each cohort. The odds ratios for BD for individuals in the top decile of PRSs compared with those in the lowest decile and middle decile were calculated in the 52 datasets internal to the PGC. To assess cross-ancestry performance, PRSs generated from the meta-analysis results were tested for association with BD using similar methods in a Japanese sample,45 a Korean sample46 and two admixed African American samples. Full details of the QC, imputation and analysis of these samples are in the Supplementary Note.

LDSC. LDSC was used to estimate the h2SNP of BD from GWAS summary statistics. MSCP was converted to the liability scale, using a lifetime BD prevalence of 2% and 1%. LDSC bivariate genetic correlations attributable to genome-wide SNPs (r2) were estimated with 255 human diseases and traits from published GWAS and 514 GWASs of phenotypes in the UK Biobank from LD Hub.47 Adjusting for the number of traits tested, the Bonferroni-corrected P-value thresholds were P < 1.96 × 10–8 and P < 9.73 × 10–9, respectively.

MixR. We applied causal mixture models to the GWAS summary statistics, using MixR v1.3. MixR provides univariate estimates of the proportion of non-null SNPs (‘polygenicity’) and the variance of effect sizes of non-null SNPs (‘discoverability’) in each phenotype. For each SNP j, univariate MixR models its additive genetic effect of allele substitution, βj as a point-normal mixture, βj = 1 − π r N (0, 0.5 + π N(0, σj2), where πj represents the proportion of non-null SNPs (‘polygenicity’) and σj2 represents variance of effect sizes of non-null SNPs (‘discoverability’). Then, for each SNP j, MixR incorporates LD information and allele frequencies for M = 9,997,231 SNPs extracted from 1000 Genomes phase 3 data to estimate the expected probability distribution of the significance statistic, ζj = βj + N(0, σj2), where ζj = βj / σj2. For a phenotype size, H, indicates heterozygosity of the jth SNP, rj indicates allelic correlation between the jth and jth SNPs and ηj = N(0, σj2) is the residual variance. Further, the three parameters, πj, σj2 and ηj2 are fitted by direct maximization of the likelihood function. The optimization is based on a set of approximately 600,000 SNPs, obtained by selecting a random set of 2,000,000 SNPs with MAF of 5% or higher, followed by LD pruning at LD r2 < 0.8 threshold. The random SNP selection and all regression to the discoverability and standard errors of model parameters. The log-likelihood figure shows individual curves for each of the 20 runs, each shifted vertically so that the best log-likelihood point is shown at the zero ordinate.

The total number of trait-influencing variants is estimated as Mr, where M = 9,997,231 gives the number of SNPs in the reference panel. MixR Venn diagram reports the effective number of influencing variants, mMr, where m is a fixed number, η = 0.319, which gives the fraction of influencing variants contributing to 90% of the trait’s heritability (with rationale for this adjustment being that the remaining 68.1% of influencing variants are small and cumulatively explain only 10% of the trait’s heritability). Phenotypic variance explained on average by an influencing genetic variant is calculated as Hr2, where H = 1/n Hrj2 is the average heterozygosity across SNPs in the reference panel. Under the assumptions of the MixR model, SNP heritability is then calculated as h2SNP = mMr × Hr2.

In the cross-trait analysis, MixR models additive genetic effects as a mixture of four components, representing null SNPs in both traits (xj); SNPs with a specific effect on the first and on the second trait (πj and πj2); and SNPs with nonzero effect on both traits (πj2). In the last component, MixR models the variance–covariance matrix as Σi,j = σj2, where the coefficients of covariation between the shared component, and σj2 and σj2 specific effect on the first and on the second trait (πj2). The fraction of influencing variants with concordant effect direction is calculated as twice the multivariate normal cumulative distribution function at point (0, 0) for the bivariate normal distribution with zero mean and variance–covariance matrix Σr. All code is available online (https://github.com/precimed/mixer).
trained interviewers, clinician-administered checklists or medical record review. In the external biobank cohorts, BD subtypes were defined using ICD codes (Supplementary Note). LDSC was used to estimate the $R^2_{LDSC}$ of each subtype, and the genetic correlation between the subtypes. The difference between the LDSC $R^2_{LDSC}$ estimates for BD I and BD II was tested for deviation from 0 using the block jackknife. The LDSC genetic correlation ($r_g$) was tested for difference from 1 by calculating a chi-square statistic corresponding to the estimated $r_g$ as $(r_g - 1)/n$.

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

**Data availability**

GWAS summary statistics are publicly available on the PGC website (https://www.med.yale.edu/pgc/results-and-downloads). Individual-level data are accessible through collaborative analysis proposals to the Bipolar Disorder Working Group of the PGC (https://www.med.yale.edu/pgc-shared-methods/how-to/). This study included some unpublished datasets accessed through dbGaP (PGC bundle phs001254) and the HRC reference panel v1.0 (http://www.haplotype-reference-consortium.org/home). Databases used: Drug–Gene Interaction Database DGGdb v2.0 (https://www.dgdb.org); Psychoactive Drug Screening Database Ki DB (https://pdsdp.ucb.university/databases/kidb.php); DrugBank 5.0 (https://www.drugbank.ca); LD Hub (http://ldc.broadinstitute.org); FUMA (https://fuma.ctglab.nl).

**Code availability**

All software used is publicly available at the URLs or references cited.

**References**

92. Moon, S. et al. The Korea Biobank Array: design and identification of coding variants associated with blood biochemical traits. *Sci. Rep.* 9, 1382 (2019).

93. Bigdeli, T. B. et al. Contributions of common genetic variants to risk of schizophrenia among individuals of African and Latino ancestry. *Mol. Psychiatry* 25, 2455–2467 (2020).

94. Lam, M. et al. RICOPLII: Rapid Imputation for CoSortias PipeLine. *Bioinformatics* 36, 930–933 (2020).

95. Price, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 51, 63–73 (2019).

96. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Genome Biol.* 17, 120 (2016).

97. Purves, K. L. et al. A major role for common genetic variation in anxiety disorders. *Mol. Psychiatry* 25, 3292–3303 (2020).

98. Yu, D. et al. Interrogating the genetic determinants of Tourette’s syndrome and other tic disorders through genome-wide association studies. *Am. J. Psychiatry* 176, 217–227 (2019).

99. Watson, H. J. et al. Genome-wide association study identifies eight risk loci and implicates metabo-psychiatric origins for anorexia nervosa. *Nat. Genet.* 52, 1279–1283 (2016).

100. Grove, J. et al. Investigating the genetic determinants of Tourette’s syndrome and other tic disorders through genome-wide association studies. *Am. J. Psychiatry* 176, 1207–1214 (2019).

101. Grove, J. et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* 51, 431–444 (2019).

102. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genome features. *Bioinformatics* 26, 841–842 (2010).

103. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Gigascience* 48, gna1279–gna1283 (2019).

104. Moon, S. et al. The Korea Biobank Array: design and identification of coding variants associated with blood biochemical traits. *Sci. Rep.* 9, 1382 (2019).

105. Price, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 51, 63–73 (2019).

106. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Genome Biol.* 17, 120 (2016).

107. Purves, K. L. et al. A major role for common genetic variation in anxiety disorders. *Mol. Psychiatry* 25, 3292–3303 (2020).

108. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Gigascience* 48, gna1279–gna1283 (2019).

109. Price, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 51, 63–73 (2019).

110. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Genome Biol.* 17, 120 (2016).

111. Grove, J. et al. Investigating the genetic determinants of Tourette’s syndrome and other tic disorders through genome-wide association studies. *Am. J. Psychiatry* 176, 217–227 (2019).

112. Watson, H. J. et al. Genome-wide association study identifies eight risk loci and implicates metabo-psychiatric origins for anorexia nervosa. *Nat. Genet.* 52, 1279–1283 (2016).

113. Grove, J. et al. Investigating the genetic determinants of Tourette’s syndrome and other tic disorders through genome-wide association studies. *Am. J. Psychiatry* 176, 1207–1214 (2019).

114. Grove, J. et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* 51, 431–444 (2019).

115. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genome features. *Bioinformatics* 26, 841–842 (2010).

116. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Gigascience* 48, gna1279–gna1283 (2019).

117. Purves, K. L. et al. A major role for common genetic variation in anxiety disorders. *Mol. Psychiatry* 25, 3292–3303 (2020).

118. Yu, D. et al. Interrogating the genetic determinants of Tourette’s syndrome and other tic disorders through genome-wide association studies. *Am. J. Psychiatry* 176, 217–227 (2019).

119. Watson, H. J. et al. Genome-wide association study identifies eight risk loci and implicates metabo-psychiatric origins for anorexia nervosa. *Nat. Genet.* 52, 1279–1283 (2016).

120. Grove, J. et al. Investigating the genetic determinants of Tourette’s syndrome and other tic disorders through genome-wide association studies. *Am. J. Psychiatry* 176, 1207–1214 (2019).

121. Seielstad, M. et al. The Korea Biobank Array: design and identification of coding variants associated with blood biochemical traits. *Sci. Rep.* 9, 1382 (2019).
R.M.M., M.M.N., M. Ribasés, G.A.R., T.S., S. Cichon. Obtained funding for BD samples: M.I., M. Cairns, I.N.F., E. Frisén, S.I., Y.K., J.A.K., M. Kubo, C. Lavebratt, S.L., D.M., P. McGuffin, G.W.M., J.R.P., M.H.R., J.R.D., D. J. Stein, J.S.S., C. Terao, A.H.Y., P.P.Z., M.A., I. Alfredsson, I.B., B.T.B., F.B., W.H.B., M. Boekhne, A.D.B., G. Breen, A.C., N.C., B.E., M.F., J.M.F., E.S.G., M. J. Green, M.G.S., K. S.H., K. Hveem, N.I., I.J., L.A.J., M. Landén, M. Leboyer, N.G.M., E.J.M., P. B. Mitchell, O.M., P.B. Mortensen, B.M.N., M.N., M.M.N., M.C.O., D., T.O., M.J.O., C. Patò, M.T.P., G.P.P., M. Rietschel, G.A.R., T.S., M.S., P.R.S., T.G.S., C.S.W., J.W.S., G.T., J.B.V., T.W.W., T.W., J.M.B., J.J.N., H.J.E., R.A.O., O.A.A. Statistical analysis: N.M., K.S.O.C., B.C., J.R.I.C., Z.Q., T.D.A., T.B.B., S. Børte, J.B., A.W.C., O.K.D., M. J. Gandal, S.P.H., N.K., M. Kim, K.K., G.P., B.M.S., L.G.S., S. Steinberg, V.T., B.S.W., H.-H.W., V.A., S.A., S.E.B., B.M.B., A.M.D., A.L.D., V.E., P., T.M.F., O.F., S.D.G., T.A.G., J. Grove, P.A.H., L.H., J.S.I., Y.K., M. Kubo, C. Lavebratt, M. Leber, P.H.I., S.H.M., A. Måås, M.M., S.A.M., S.E.M., L.M.O.L., A.P., T.R., P.R., D.M.R., O.B.S., C. Terao, T.E.T., T.R.W., W.X., J.M.K.Y., H.Y., P.P.Z., H.Z., A.D.B., G. Breen, E.S.G., F.S.G., Q.S.L., B.M.M., C.M.N., D.P., S.R., H.S., P.F.S., T.W., N.R.W., J.M.B., E.A.S., K.S.O.C., B.C., J.R.I.C. and Z.Q. contributed equally to this work and should be regarded as joint second authors.

Competing interests
T.E.T., S. Steinberg, H.S. and K.S. are employed by deCODE Genetics/Amgen. Multiple additional authors work for pharmaceutical or biotechnology companies in a manner directly analogous to academic coauthors and collaborators. A.H.Y. has given paid lectures and served on advisory boards relating to drugs used in affective and related disorders for several companies (AstraZeneca, Eli Lilly, Lundbeck, Sunovion, Servier, Livanova, Janssen, Allergan, Bionomics and Sumitomo Dainippon Pharma), was Lead Investigator for Embolden Study (AstraZeneca), BCI Neuropsychiatry study and Aripiprazole Mania Study, and is an investigator for Janssen, Lundbeck, Livanova and Compass. J.I.N. is an investigator for Janssen. P.E.S. reports the following potentially competing financial interests: Lundbeck (advisory committee), Pfizer (Scientific Advisory Board member) and Roche (grant recipient, speaker reimbursement). G. Breen reports consultancy and speaker fees from Eli Lilly and Illumina and grant funding from Eli Lilly. M. Landén has received speaker fees from Lundbeck. O.A.A. has received speaker fees from Lundbeck and Sunovion, and is a consultant to HealthLytx. J.A.R.-Q. was on the speakers bureau and/or acted as consultant for Eli Lilly. Janssen-Cilag, Novartis, Shire, Lundbeck, Almirall, Braingaze, Sinclairlab and Rubió in the last 5 years. He also received travel awards (air tickets and hotel) for taking part in psychiatric meetings from Janssen-Cilag, Rubió, Shire and Eli Lilly. The Department of Psychiatry chaired by him received unrestricted educational and research support from the following companies in the last 5 years: Eli Lilly, Lundbeck, Janssen-Cilag, Actelion, Shire, Ferrer, Oryzon, Roche, Pistoia and Rubió. E.V. has received grants and served as a consultant, advisor or CME speaker for the following entities: AB-Biotics, Abbott, Allergan, Angelini, AstraZeneca, Bristol Myers Squibb, Dainippon Sumitomo Pharma, Farindustria, Ferrer, Forest Research Institute, Gedeon Richter, GlaxoSmithKline, Janssen, Lundbeck, Otsuka, Pfizer, Roche, SAGE, Sanofi-Aventis, Servier, Shire, Sunovion, Takeda, the Brain and Behaviour Foundation, the Catalan Government (AGAUR and PERIS), the Spanish Ministry of Science, Innovation, and Universities (AES and CIBERSAM), the Seventh European Framework Programme and Horizon 2020 and the Stanley Medical Research Institute. T. Elvsåshagen has received speaker fees from Lundbeck. S.K.-S. received author's and consultant honoraria from Medice Arzneimittel Pütter GmbH and Shire/Takeda. A.S. is or has been a consultant/speaker for: Abbott, Abbvie, Angelini, AstraZeneca, Clinical Data, Boehringer, Bristol Myers Squibb, Eli Lilly, GlaxoSmithKline, Innopavarma, Italfarmaco, Janssen, Lundbeck, Naurex, Pfizer, Pelifarma, Sanofi, Servier. J.R.D. has served as an unpaid consultant to Myriad - Neuroscience (formerly Assurex Health) in 2017 and 2019 and owns stock in CVS Health. H.R.K. serves as an advisory board member for Dicerna Pharmaceuticals, and is a member of the American Society of Clinical Psychopharmacology's Alcohol Clinical Trials Initiative, which was sponsored in the past 3 years by AbbVie, Alkermes, Amygdala Neurosciences, Arbor Pharmaceuticals, Ethypharm, Indivior, Lilly, Lundbeck, Otsuka and Pfizer. H.R.K. is named as an inventor on PCT patent application no. 15/878,640 entitled: Genotype-guided dosing of opioid agonists, filed January 24, 2018. B.M.N. is a member of the scientific advisory board at Deep Genomics and consultant for Camp4 Therapeutics, Takeda Pharmaceutical and Biogen. All other authors declare no financial interests or potential conflicts of interest.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-021-00857-4.
Correspondence and requests for materials should be addressed to N.M. or O.A.A.
Peer review information Nature Genetics thanks Na Cai, Qiang Wang and the other anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.
Reprints and permissions information is available at www.nature.com/reprints.
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement.
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly.
- The statistical test(s) used AND whether they are one- or two-sided.
- Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested.
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons.
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals).
- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted.
- Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings.
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes.
- Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated.

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection: Ascertainment of all cohorts is described in the Supplementary Note. Data were generated at many sites using standard genotyping pipelines and calling softwares from commercial sources (Affymetrix and Illumina).

Data analysis: Analyses were performed using the publicly available RICOPIL software suite [version 2018_Nov_23.001] which provides wrappers for standard genetic analysis software including PLINK v1.09, Minimac3, Eagle v2.3.5, EIGENSTRAT v6.1.4 and METAL [version 2011-03-25]. Post-GWAS analyses were conducted using FUMA v1.3.6a, MAGMA v1.08, FUSION (vOct 1, 2019), FOCUS 0.6.10, SMR v1.03, LDSC v1.0.0, MlioR v1.3, GEMR implemented in GCTA software v1.93.11 beta, bedtools v2.29.2, TwoSampleMR v3.5.4 and MR-PRESSO v1.0 R packages. All software is publicly available. Any additional analysis code is detailed in Methods and the Supplementary Note.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

GWAS summary statistics are publicly available on the PGC website (https://www.med.unc.edu/pgc/results-and-downloads). Individual-level data are accessible through collaborative analysis proposals to the Bipolar Disorder Working Group of the PGC (https://www.med.unc.edu/pgc/shared-methods/how-to/). This study included some publicly available datasets accessed through dbGaP (PGC bundle phs001254.v1.p1) and the Haploview Reference Consortium reference panel v1.0.
Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- [ ] Life sciences
- [ ] Behavioural & social sciences
- [ ] Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

Sample size was not predetermined but all available cohorts of bipolar disorder cases and controls were included to maximize sample size. This study provides a 1.5-fold increase in effective sample size compared with previous GWAS of bipolar disorder, increasing statistical power to detect novel genetic associations.

Data exclusions

Predetermined phenotypic data exclusions, for both cases and controls, are detailed in the Supplementary Note. Genotype data exclusions were also predetermined and were performed for quality control; these included high missing call rate, high or low heterozygosity, inconsistent genotype versus clinical data sex, and ancestry outlier status based on visual inspection of genotype principal component analysis results. Phenotypic and genotypic exclusions were applied to the GWAS and all post-GWAS analyses.

Replication

All available cohorts of bipolar disorder cases and controls were included in the primary analysis and therefore we do not perform replication of genetic associations in independent cohorts here. Post-GWAS analyses were replicated using different statistical genetics methods or by integrating different publicly available biological datasets where possible.

Randomization

Samples/participants were allocated into experimental groups by clinical cohort (which included country of origin) and genotype data collection batches. Association analyses were performed in each dataset and meta-analyzed across datasets. Ancestry covariates derived from genotype principal components analysis were included in association tests, which were logistic regression.

Blinding

Case and control groups were assigned by trained researchers or interviewers during cohort ascertainment, or using predefined International Classification of Diseases codes applied to biobanks. Therefore case/control status was fixed at ascertainment. Experimenters were not blind to case/control status while performing statistical analyses of the genetic data as this was not practical.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

| Materials & experimental systems | Methods |
|----------------------------------|---------|
| n/a                             | n/a     |
| [ ] Involved in the study        | Involved in the study |
| [ ] Antibodies                  | [ ] ChiP-seq |
| [ ] Eukaryotic cell lines        | [ ] Flow cytometry |
| [ ] Palaeontology and archaeology| [ ] MRI-based neuroimaging |
| [ ] Animals and other organisms |         |
| [ ] Human research participants |         |
| [ ] Clinical data               |         |
| [ ] Dual use research of concern|         |

Human research participants

Policy information about studies involving human research participants

Population characteristics

This is fully described in the Online Methods and associated Supplementary Note and Tables.

Recruitment

This is fully described in the Online Methods and associated Supplementary Note and Tables.

Ethics oversight

All local IRBs approved of this study. This is fully described in the Online Methods and associated Supplementary Note and Tables.

Note that full information on the approval of the study protocol must also be provided in the manuscript.