Fine-mapping genomic loci refines bipolar disorder risk genes

Authors and Affiliations

Authors

Maria Koromina[1,2,3,*], Ashvin Ravi[3,4,5,6], Georgia Panagiotaropoulou[7], Brian M. Schilder[3,5,6], Jack Humphrey[3,4,5,6], Alice Braun[7], Tim Bidgeli[8], Chris Chatzinakos[8], Brandon Coombes[9], Jaeyoung Kim[10,11], Xiaoxi Liu[12,13], Chikashi Terao[12,13,14], Kevin S. O.'Connell[15,16], Mark Adams[17], Rolf Adolfsson[18], Martin Alda[19,20], Lars Alfredsson[21], Till F. M. Andlauer[22], Ole A. Andreassen[15,16], Anastasia Antoniou[23], Bernhard T. Baune[24,25,26], Susanne Bengesser[27], Joanna Biernacka[28,29], Michael Boehnke[30], Rosa Bosch[31,32], Murray Cairns[33], Vaughan J. Carr[34], Miquel Casas[31,32], Stanley Catts[35], Sven Cichon[36,37,38,39], Aiden Corvin[40], Nicholas Craddock[41], Konstantinos Dafnas[23], Nina Dalkner[27], Udo Dannlowski[42], Franziska Degenhardt[37,43], Arianna Di Florio[41,44], Dimitris Dikeos[23], Frederike Tabea Fellendorf[27], Panagiotis Ferentinos[23,45], Andreas J. Forstner[37,39,46], Liz Forty[41], Mark Frye[29], Janice M. Fullerton[47,48], Micha Gawlik[49], Ian R. Gizer[50], Katherine Gordon-Smith[51], Melissa J. Green[47,52], Maria Grigoroiu-Serbanescu[53], José Guzman-Parra[54], Tim Hahn[42], Frans Henskens[33], Jan Hillert[55], Assen V. Jablensky[56], Lisa Jones[51], Ian Jones[41], Lina Jonsson[57], John R. Kelsoe[58], Tilo Kircher[59], George Kirov[41], Sarah Kittel-Schneider[49,60,61], Manolis Kogevinas[62], Mikael Landén[63,63], Marion Leboyer[64,65], Melanie Lenger[27], Jolanta Lissowska[66], Christine Lochner[67], Carmel Loughland[33], Donald MacIntyre[17], Nicholas G. Martin[68,69], Eirini Maratou[70], Carol A. Mathews[71], Fermin Mayoral[54], Susan L. McElroy[72], Nathaniel W. McGregor[73], Andrew McIntosh[17], Andrew McQuillin[74], Patricia Michie[33], Vihra Milanova[75], Philip B. Mitchell[52], Paraskevi Moutsatsou[76], Bryan Mowry[35], Bertram Müller-Myhsok[77,78], Richard Myers[79], Igor Nenadić[80,81], Markus M. Nöthen[37], Claire O'Donovan[19], Michael O'Donovan[41], Roel A. Ophoff[82,83,84,85], Michael J Owen[41], Chris Pantelis[86], Carlos Pato[87], Michele T. Pato[87], George P. Patrinos[88,89,90], Joanna M. Pawlak[91], Roy H. Perlis[92,93], Evgenia Porichi[94], Danielle Posthuma[95,96], Josep Antoni Ramos-Quiroga[31,97,98,99], Andreas Reif[60], Eva Z. Reininghaus[27], Marta Ribasés[31,100,101,102], Marcella Rietschel[103], Ulrich Schall[33], Thomas G. Schulze[103,104,105,106,107], Laura Scott[30], Rodney J. Scott[33], Alessandro Serretti[108], Cynthia Shannon Weickert[52,109], Jordan W. Smoller[110,111,112], Maria Soler Artigas[31,97,99,113], Dan J. Stein[114], Fabian Streit[103], Claudio Toma[47,52,115], Paul Tooney[33], Eduard Vieta[116], John B. Vincent[117], Irwin D. Waldman[118], Thomas Weickert[52,109], Stephanie H. Witt[103], Kyung Sue Hong[119], Masashi Ikeda[120], Nakao Iwata[120], Beata Świątkowska[121], Hong-Hee Won[10], Howard J. Edenberg[122,123], Stephan Ripke[7,124], Towfique Raj[3,5,6], Jonathan R. I. Coleman[45,125], Niamh Mullins[1,2,3,*]

Affiliations

[1]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [2]Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [3]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [4]Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [5]Ronald M. Loeb Center for Alzheimer's Disease, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [6]Estelle and Daniel Maggin Department of Neurology, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [7]Department of Psychiatry and Psychotherapy, Charité - Universitätsmedizin, Berlin, Germany. [8]SUNY Downstate Health Sciences University. [9]Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA. [10]Samsung Advanced Institute for Health Sciences and Technology (SAIHST), Sungkyunkwan University, Samsung Medical Center, Seoul, Republic of Korea. [11]Samsung Genome Institute, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea. [12]Laboratory for Statistical and Translational Genetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. [13]Clinical Research Center, Shizuoka General Hospital, Shizuoka, Japan. [14]The Department of Applied Genetics, The School of Pharmaceutical Sciences, University of Shizuoka, Shizuoka, Japan.

[15]Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway. [16]NORMENT, University of Oslo, Oslo, Norway. [17]Division of Psychiatry, Centre for Clinical Brain Sciences, The University of Edinburgh, Edinburgh, UK. [18]Department of Clinical Sciences, Psychiatry, Umeå, University Medical Faculty, Umeå, Sweden. [19]Department of Psychiatry, Dalhousie University, Halifax, NS, Canada. [20]National Institute of Mental Health, Klecany, Czech Republic. [21]Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden. [22]Department of Neurology, Klinikum rechts der Isar, School of Medicine, Technical University of Munich, Munich, Germany. [23]National Kapodistrian University of Athens, 2nd Department of Psychiatry, Attikon General Hospital, Athens, Greece. [24]Department of Psychiatry, University of Münster, Münster, Germany. [25]Department of Psychiatry, Melbourne Medical School, The University of Melbourne, Melbourne, VIC, Australia. [26]The Florey Institute of Neuroscience and Mental Health, The University of Melbourne, Parkville, VIC, Australia. [27]Medical University of Graz, Division of Psychiatry and Psychotherapeutic Medicine, Graz, Austria. [28]Department of Quantitative Health Sciences Research, Mayo Clinic, Rochester, MN, USA. [29]Department of Psychiatry and Psychology, Mayo Clinic, Rochester, MN, USA. [30]Center for Statistical Genetics and Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA. [31]Instituto de Salud Carlos III, Biomedical Network Research Centre on Mental Health (CIBERSAM), Madrid, Spain. [32]Programa SJD MIND Escoles, Hospital Sant Joan de Déu, Institut de Recerca Sant Joan de Déu, Esplugues de Llobregat, Spain. [33]University of Newcastle, Newcastle, NSW, Australia. [34]School of Clinical Medicine, Discipline of Psychiatry and Mental Health, University of New South Wales, Sydney, NSW, Australia. [35]University of Queensland, Brisbane, QLD, Australia. [36]Department of Biomedicine, University of Basel, Basel, Switzerland. [37]Institute of Human Genetics, University of Bonn, School of Medicine and University Hospital Bonn, Bonn, Germany. [38]Institute of Medical Genetics and Pathology, University Hospital Basel, Basel, Switzerland. [39]Institute of Neuroscience and Medicine (INM-1), Research Centre Jülich, Jülich, Germany. [40]Neuropsychiatric Genetics Research Group, Dept of Psychiatry and Trinity Translational Medicine Institute, Trinity College Dublin, Dublin, Ireland. [41]Centre for Neuropsychiatric Genetics and Genomics, Division of Psychological Medicine and Clinical Neurosciences, Cardiff University, Cardiff, UK. [42]Institute for Translatiol Psychiatry, University of Münster, Münster, Germany. [43]Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University Hospital Essen, University of Duisburg-Essen, Duisburg, Germany. [44]Department of Psychiatry, University of North Caroli at Chapel Hill, Chapel Hill, NC, USA. [45]Social, Genetic and Developmental Psychiatry Centre, King's College London, London, UK. [46]Centre for Human Genetics, University of Marburg, Marburg, Germany. [47]Neuroscience Research Australia, Sydney, NSW, Australia. [48]School of Biomedical Sciences, University of New South Wales, Sydney, NSW, Australia. [49]Department of Psychiatry, Psychosomatics and Psychotherapy, Center of Mental Health, University Hospital Würzburg, Würzburg, Germany. [50]Department of Psychological Sciences, University of Missouri, Columbia, MO, USA. [51]Psychological Medicine, University of Worcester, Worcester, UK. [52]Discipline of Psychiatry and Mental Health, School of Clinical Medicine, Faculty of Medicine and Health, University of New South Wales, Sydney, NSW, Australia. [53]Biometric Psychiatric Genetics Research Unit, Alexandru Obregia Clinical Psychiatric Hospital, Bucharest, Romania. [54]Mental Health Department, University Regional Hospital, Biomedicine Institute (IBIMA), Málaga, Spain. [55]Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden. [56]University of Western Australia, Nedlands, WA, Australia. [57]Institute of Neuroscience and Physiology, University of Gothenburg, Gothenburg, Sweden. [58]Department of Psychiatry, University of California San Diego, La Jolla, CA, USA. [59]Department of Psychiatry and Psychotherapy, University of Marburg, Germany. [60]Department of Psychiatry, Psychosomatic Medicine and Psychotherapy, University Hospital Frankfurt, Frankfurt am Main, Germany. [61]Department of Psychiatry and Neurobehavioural Science, University College Cork, Cork, Ireland. [62]ISGlobal, Barcelona, Spain. [63]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [64]Université Paris Est Créteil, INSERM, IMRB, Translational Neuropsychiatry, Créteil, France. [65]Department of Psychiatry and Addiction Medicine, Assistance Publique - Hôpitaux de Paris, Paris, France. [66]Cancer Epidemiology and Prevention, M. Sklodowska-Curie National Research Institute of Oncology, Warsaw, Poland. [67]SA MRC Unit on Risk and Resilience in Mental Disorders, Dept of Psychiatry, Stellenbosch University, Stellenbosch, South Africa. [68]Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia. [69]School of Psychology, The University of Queensland, Brisbane, QLD, Australia. [70]National and Kapodistrian University of Athens, Medical School, Clinical Biochemistry Laboratory, Attikon General Hospital, Athens, Greece. [71]Department of Psychiatry and Genetics Institute, University of Florida, Gainesville, FL, USA. [72]Research Institute, Lindner Center of HOPE, Mason, OH, USA. [73]Systems Genetics Working Group, Department of Genetics, Stellenbosch University, Stellenbosch, South Africa. [74]Division of Psychiatry, University College London, London, UK. [75]Psychiatric Clinic, Alexander University Hospital, Bulgaria. [76]National Kapodistrian University of

Athens, Medical School, Clinical Biochemistry Laboratory, Attikon General Hospital, Athens, Greece. [77]Department of Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Munich, Germany. [78]Munich Cluster for Systems Neurology (SyNergy), Munich, Germany. [79]Hudsolpha Institute for Biotechnology, Huntsville, AL, USA. [80]Department of Psychiatry and Psychotherapy, University of Marburg, Marburg, Germany. [81]Center for Mind, Brain and Behavior (CMBB), University of Marburg and Justus Liebig University Giessen, Giessen, Germany. [82]Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, Los Angeles, CA, USA. [83]Department of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA. [84]Department of Psychiatry, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands. [85]Department of Psychiatry and Biobehavioral Science, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA. [86]University of Melbourne, VIC, Australia. [87]Institute for Genomic Health, SUNY Downstate Medical Center College of Medicine, Brooklyn, NY, USA. [88]University of Patras, School of Health Sciences, Department of Pharmacy, Laboratory of Pharmacogenomics and Individualized Therapy, Patras, Greece. [89]United Arab Emirates University, College of Medicine and Health Sciences, Department of Genetics and Genomics, Al-Ain, United Arab Emirates. [90]United Arab Emirates University, Zayed Center for Health Sciences, Al-Ain, United Arab Emirates. [91]Department of Psychiatry, Departmet of Psychiatric Genetics, Poznan University of Medical Sciences, Poznan, Poland. [92]Psychiatry, Harvard Medical School, Boston, MA, USA. [93]Division of Clinical Research, Massachusetts General Hospital, Boston, MA, USA. [94]National and Kapodistrian University of Athens, 2nd Department of Psychiatry, Attikon General Hospital, Athens, Greece. [95]Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. [96]Department of Clinical Genetics, Amsterdam Neuroscience, Vrije Universiteit Medical Center, Amsterdam, The Netherlands. [97]Department of Psychiatry, Hospital Universitari Vall d´Hebron, Barcelo, Spain. [98]Department of Psychiatry and Forensic Medicine, Universitat Autònoma de Barcelo, Barcelo, Spain. [99]Psychiatric Genetics Unit, Group of Psychiatry Mental Health and Addictions, Vall d´Hebron Research Institut (VHIR), Universitat Autònoma de Barcelo, Barcelo, Spain. [100]Department of Psychiatry, Hospital Universitari Vall d´Hebron, Barcelona, Spain. [101]Psychiatric Genetics Unit, Group of Psychiatry Mental Health and Addictions, Vall d´Hebron Research Institut (VHIR), Universitat Autònoma de Barcelona, Barcelona, Spain. [102]Department of Genetics, Microbiology, and Statistics, Faculty of Biology, Universitat de Barcelona, Barcelona, Spain.Department of Genetics, Microbiology and Statistics, Faculty of Biology, Universitat de Barcelona, Barcelona, Catalonia, Spain. [103]Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany. [104]Institute of Psychiatric Phenomics and Genomics (IPPG), University Hospital, LMU Munich, Munich, Germany. [105]Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [106]Department of Psychiatry and Psychotherapy, University Medical Center Göttingen, Göttingen, Germany. [107]Department of Psychiatry and Behavioral Sciences, SUNY Upstate Medical University, Syracuse, NY, USA. [108]Department of Biomedical and NeuroMotor Sciences, University of Bologna, Bologna, Italy. [109]Department of Neuroscience, SUNY Upstate Medical University, Syracuse, NY, USA. [110]Stanley Center for Psychiatric Research, Broad Institute, Cambridge, MA, USA. [111]Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA. [112]Psychiatric and Neurodevelopmental Genetics Unit (PNGU), Massachusetts General Hospital, Boston, MA, USA. [113]Department of Genetics, Microbiology, and Statistics, Faculty of Biology, Universitat de Barcelo, Barcelo, Spain. [114]SAMRC Unit on Risk and Resilience in Mental Disorders, Dept of Psychiatry and Neuroscience Institute, University of Cape Town, Cape Town, South Africa. [115]Centro de Biología Molecular Severo Ochoa, Universidad Autónoma de Madrid and CSIC, Madrid, Spain. [116]Clinical Institute of Neuroscience, Hospital Clinic, University of Barcelona, IDIBAPS, CIBERSAM, Barcelona, Spain. [117]Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, ON, Canada. [118]Department of Psychology, Emory University, Atlanta, GA, USA. [119]Department of Psychiatry, Sungkyunkwan University School of Medicine, Samsung Medical Center, Seoul, South Korea. [120]Department of Psychiatry, Fujita Health University School of Medicine, Toyoake, Japan. [121]Department of Environmental Epidemiology, Nofer Institute of Occupational Medicine, Lodz, Poland. [122]Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN, USA. [123]Department of Medical and Molecular Genetics, Indiana University, Indianapolis, IN, USA. [124]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA. [125]NIHR Maudsley Biomedical Research Centre, South London and Maudsley NHS Foundation Trust, London, UK.

*Corresponding authors: Maria Koromina, maria.koromina@mssm.edu and Niamh Mullins, niamh.mullins@mssm.edu

Abstract

Bipolar disorder (BD) is a heritable mental illness with complex etiology. While the largest published genome-wide association study identified 64 BD risk loci, the causal SNPs and genes within these loci remain unknown. We applied a suite of statistical and functional fine-mapping methods to these loci, and prioritized 22 likely causal SNPs for BD. We mapped these SNPs to genes, and investigated their likely functional consequences by integrating variant annotations, brain cell-type epigenomic annotations, brain quantitative trait loci, and results from rare variant exome sequencing in BD. Convergent lines of evidence supported the roles of *SCN2A, TRANK1, DCLK3, INSYN2B, SYNE1, THSD7A, CACNA1B, TUBBP5, PLCB3, PRDX5, KCNK4, AP001453.3, TRPT1, FKBP2, DNAJC4, RASGRP1, FURIN, FES, YWHAE, DPH1, GSDMB, MED24, THRA, EEF1A2*, and *KCNQ2* in BD. These represent promising candidates for functional experiments to understand biological mechanisms and therapeutic potential. Additionally, we demonstrated that fine-mapping effect sizes can improve performance and transferability of BD polygenic risk scores across ancestrally diverse populations, and present a high-throughput fine-mapping pipeline (https://github.com/mkoromina/SAFFARI).

Introduction

Bipolar disorder (BD) is a heritable mental illness with complex etiology[1]. Heritability estimates from twin studies range between 60% and 90%[2–4], while SNP-based heritability (h^2_{SNP}) calculations suggest that common genetic variants can explain up to 20% of the phenotypic variance of BD[5]. Genome-wide association studies (GWAS) of common variants have been successful in identifying associated genetic risk loci for BD[5–15]. For example, the largest published BD GWAS to date, conducted by the Psychiatric Genomics Consortium (PGC), comprised more than 40,000 BD cases and 370,000 controls from 57 cohorts of European ancestries, and identified 64 genome-wide significant (GWS) risk loci[16]. However, identifying the causal SNPs within these loci (i.e., SNPs responsible for the association signal at a locus and with a biological effect on the phenotype, as opposed to those associated due to linkage disequilibrium (LD) with a causal variant) is a major challenge.

Computational fine-mapping methods aim to identify independent causal variants within a genomic locus by modeling LD structure, SNP association statistics, number of causal variants, and/or prior probabilities of causality based on functional annotations. There are a variety of fine-mapping models ranging from regression to Bayesian methods, with different strengths and limitations[17–19]. For example, the Sum of Single Effects (SuSiE) model uses iterative

Bayesian selection with posterior probabilities[20], FINEMAP employs a stochastic search algorithm for SNP combinations[21], and POLYgenic FUNctionally-informed fine-mapping (PolyFun) computes functional priors to improve fine-mapping accuracy[18]. Bayesian fine-mapping methods typically generate a posterior inclusion probability (PIP) of causality per SNP, and "credible sets" of SNPs, which represent the minimum set of SNPs with a specified probability of including the causal variant(s). Many methods can assume one or multiple causal variants per locus, and can now be applied to GWAS summary statistics from large and well-powered studies. This is highly advantageous for fine-mapping GWAS meta-analyses; however, the specification of appropriate LD structure is crucial for accurate fine-mapping. When LD cannot be obtained from the original cohort(s) (e.g. due to data access restrictions), it should instead be obtained from a sufficiently large sample that is ancestrally similar to the GWAS population[22].

Fine-mapping methods have recently been applied to GWAS of psychiatric disorders. For example, a recent study using FINEMAP and integrating functional genomic data identified more than 100 genes likely to underpin associations in risk loci for schizophrenia[23]. Several fine-mapped candidates had particularly strong support for their pathogenic role in schizophrenia, due to convergence with rare variant associations[23]. Here, we use a suite of tools to conduct statistical and functional fine-mapping of 64 GWS risk loci for BD[16] and assess the impact of the LD reference panel and fine-mapping window specifications. We link the likely causal SNPs to their relevant genes and investigate their potential functional consequences, by integrating functional genomic data, including brain cell-type-specific epigenomic annotations, and quantitative trait loci data. We also fine-mapped the major histocompatibility complex (MHC) separately by imputing human leukocyte antigen (HLA) variants, and assessed the impact of fine-mapping on polygenic risk score (PRS) predictions. Finally, we present a comprehensive fine-mapping pipeline implemented via Snakemake[24] as a rapid, scalable, and cost-effective approach to prioritize likely variants from GWS risk loci. This strategy yielded promising candidate genes for future experiments to understand the mechanisms by which they increase risk of BD.

Methods

GWAS summary statistics and BD risk loci

Summary statistics from the latest published BD GWAS by the Psychiatric Genomics Consortium (PGC) were used as input to the fine-mapping pipeline[16]. Briefly, this GWAS comprised 41,917 BD cases, and 371,549 controls of European (EUR) ancestries, from 57 cohorts (**Supplementary Table S1**). Most genomic data were imputed using the Haplotype Reference Consortium (HRC) EUR ancestry reference panel v1.0[25], leading to a total of 7,608,183 SNPs that were well-

imputed and well-represented across cohorts in the GWAS meta-analysis. Each GWS locus window was established around the GWS significant "top lead" SNP ($P < 5 \times 10^{-8}$), with boundaries defined by the positions of the 3'-most and 5'-most SNPs, requiring an LD $r^2 > 0.1$ with the top lead SNP within a 3 Mb range, according to the LD structure of the HRC EUR reference panel[16]. The GWAS meta-analysis identified 64 independent loci associated with BD at GWS, which were selected for fine-mapping. Due to the complexity and long-range LD of the MHC/HLA region, this locus was analyzed separately (see section 'Fine-mapping the MHC locus'). **Supplementary Table S2** shows the top lead SNP from each GWS locus, association statistics, locus boundaries, locus size, and locus names (as defined in the original GWAS)[16].

Conditional analysis

Figure 1 shows an overview of the fine-mapping pipeline. First, stepwise conditional analyses were conducted using GCTA-COJO[26] to identify potential independent association signals within each locus, according to the LD structure of the HRC EUR reference panel. Association tests were performed for all SNPs in each locus, conditioning sequentially on the top lead SNP, until no conditional tests were significant (conditional $P > 5 \times 10^{-6}$), to calculate the number of independent association signals per locus. A less stringent P value threshold ($P < 5 \times 10^{-6}$) for significance was selected for the conditional analysis, consistent with the recommendations of Yang et al[26].

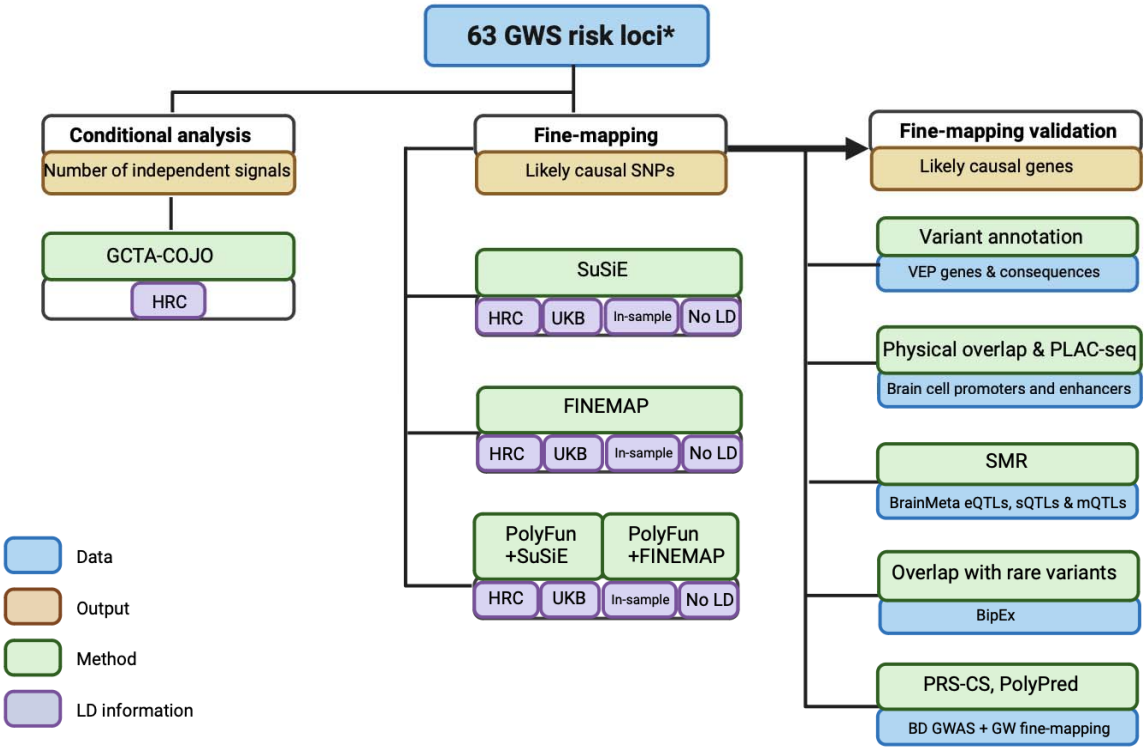


Figure 1. Schematic workflow of the fine-mapping pipeline developed for BD GWAS risk loci. Conditional analyses were performed within GWS loci using GCTA-COJO, based on the linkage disequilibrium (LD) structure of the Haplotype Reference Consortium (HRC) reference panel. Fine-mapping was conducted using statistical (SuSiE and FINEMAP) and functionally-informed (PolyFun) methods, according to the LD structure of the HRC, UK Biobank (UKB), and a subset of the GWAS data ("in-sample LD"), as well as using no LD and assuming one causal variant. PolyFun functional priors were based on the published baseline-LF2.2 UKB model[27]. Fine-mapping results were validated computationally via Variant Effect Predictor (VEP) annotations and functional consequences, overlap with epigenomic peaks from brain cell-types, Summary-data-based Mendelian Randomization analysis (SMR) with brain expression, splicing and methylation QTL data, convergence with rare variant associations from the Bipolar Exome Sequencing Collaboration (BipEx), and testing whether fine-mapping effect sizes improve polygenic risk scores (PRS-CS and PolyPred). *The major histocompatibility complex (MHC) was fine-mapped using separate procedures (see section 'Fine-mapping the MHC locus').

LD reference panels

Statistical and functional fine-mapping methods require information on LD between variants and selection of a genomic region ("window") to fine-map. To examine the impact of LD on fine-mapping, analyses were performed using LD information from the HRC EUR reference panel, published LD matrices based on EUR ancestry individuals in the UK Biobank (UKB)[18], and "in-sample" LD calculated from a subset of 48 BD cohorts in the PGC BD GWAS for which individual-level genetic data were available within the PGC (33,827 cases, 53,953 controls, all of

EUR ancestries), representing 73% of the total effective sample size of the GWAS. Briefly, HRC-imputed dosage data were converted to hard calls with a genotype call probability cut-off of 0.8 and PLINK binary files were merged across cohorts, restricting to the set of unrelated individuals included in the GWAS, using PLINK v1.90[28]. Missingness rates per SNP were calculated in each cohort, and SNPs absent in all individuals from any one cohort were excluded from the merged dataset, yielding 7,603,435 SNPs overlapping with the GWAS summary statistics. Individual-level genetic data per chromosome were used as an "in-sample" LD reference panel for fine-mapping.

Two fine-mapping "windows" were selected: 1) a 3 Mb window and 2) the GWS locus window defined in the PGC BD GWAS[16]. The 3 Mb window is a 3 Mb-long region which includes the top lead SNP of the GWS locus, predefined according to published LD matrices of 3 Mb blocks covering the entire genome calculated in the UKB[18]. The GWS locus window is established around a GWS significant "top lead" SNP ($P < 5 \times 10^{-8}$). Its boundaries are defined by the positions of the 3'-most and 5'-most SNPs, requiring an LD r^2 greater than 0.1 with the top lead SNP within a 3 Mb range, in accordance with the LD structure of the HRC EUR reference panel. Excluding the MHC, GWS locus windows ranged between 14,960 - 3,730,000 bp in size (**Supplementary Table S2**).

Statistical and functional fine-mapping

GWS loci were fine-mapped using a suite of Bayesian fine-mapping methods that can be applied to GWAS summary statistics: SuSiE, FINEMAP, PolyFun+SuSiE, PolyFun+FINEMAP (**Figure 1**). SuSiE and FINEMAP are statistical fine-mapping methods, while PolyFun incorporates functional annotations as prior probabilities to improve subsequent fine-mapping accuracy[18,20,21]. Since these methods have different underlying assumptions, strengths and limitations, results were compared to examine convergence of evidence across methods. Briefly, each Bayesian method generates SNP-wise posterior inclusion probabilities of causality (PIP), and a 95% credible set (95% CS), defined as the minimum subset of SNPs that cumulatively have at least 95% probability of containing the causal SNP(s). PIP refers to the marginal probability that a SNP is included in any causal model, conditional on the observed data, hence providing weight of evidence that a SNP should be considered potentially causal.

First, single variant fine-mapping, which makes the simple assumption of one causal variant per locus (K = 1) and does not require LD information[18,20,21], was performed using each of the four methods and both the 3 Mb and GWS locus fine-mapping windows. FINEMAP and SuSiE can assume multiple causal variants per locus, modeling the LD structure between them. Fine-mapping was additionally performed within each of the two windows, assuming the default

maximum of five causal variants per locus (K = 5) and separately using the HRC, UKB and "in-sample" LD structures. Finally, PolyFun was used to incorporate 187 published functional annotations from the baseline-LF 2.2.UKB model[27] to compute prior causal probabilities (priors) via an L2-regularized extension of stratified LD-score regression (S-LDSC)[29], and subsequently perform fine-mapping using FINEMAP and SuSiE[18]. Briefly, functional annotations included epigenomic and genomic annotations, minor allele frequency (MAF) bins, binary or continuous functional annotations, LD-related annotations such as LD level, predicted allele age, recombination rate, and CpG content[27]. Functionally-informed fine-mapping was also performed using the three LD reference panels and two fine-mapping windows.

In total, 32 fine-mapping analyses were conducted (24 multi-variant analyses using four fine-mapping methods, three LD reference panels and two fine-mapping windows, and eight LD-independent single-variant fine-mapping analyses), varying parameters to examine their impact and the convergence of results. "Consensus SNPs" were defined as those in the 95% CS from at least two methods that used the same LD structure and fine-mapping window, and with a PIP >0.95 or >0.50 (**Table 1**) (36 opportunities for a SNP to be a consensus SNP). When single-variant fine-mapping was performed, consensus SNPs were defined as those with a PIP >0.95 or PIP >0.50 and found in at least two methods using the same fine-mapping window. The "union consensus" set of SNPs was defined as all consensus SNPs across LD structure and fine-mapping windows with PIP >0.50, excluding SNPs identified only with the UKB LD reference panel and with a high GWAS P value.

All steps of the statistical and functional fine-mapping analyses have been compiled into a high-throughput pipeline named SAFFARI (**S**tatistical **A**nd **F**unctional **F**ine-mapping **A**pplied to GWAS **R**isk Loc**I**). SAFFARI is implemented through Snakemake in a Linux environment[24], with options to provide sets of GWAS summary statistics, lists of fine-mapping windows, and to specify LD reference panels, in the form of LD matrices or individual-level genetic data (GitHub: https://github.com/mkoromina/SAFFARI).

Annotation of union consensus SNPs

Union consensus SNPs were characterized using the Variant Effect Predictor (VEP) (GRCh37) Ensembl release 109[30]. When SNPs were mapped to multiple transcripts, the most severe variant consequence was retained for annotation, and when SNPs fell within intergenic or regulatory regions, no genes were annotated[30]. If annotated genes overlapped and the SNP had the same severity consequence, then both genes were annotated. Additional annotations included the CADD scores (https://cadd.gs.washington.edu/), which denote the likelihood of the variant being deleterious or disease-causing (CADD >= 20) and ClinVar annotations (https://www.ncbi.nlm.nih.gov/clinvar/) describing the association of variants with diseases

(i.e., benign, pathogenic, etc). Union consensus SNPs were further annotated with RegulomeDB (v.2.2) to determine whether they have functional consequences and lie in non-coding regions and to annotate them to the relevant regulatory elements[31]. RegulomeDB probability and ranking scores are positively correlated and predict functional variants in regulatory elements. Probability scores closer to 1 and ranking scores below 2 provide increased evidence of a variant to be in a functional region[31]. Probability of being loss-of-function intolerant (pLI) and loss-of-function observed/expected upper bound fraction (LOEUF) scores were retrieved from the Genome Aggregation Database (gnomAD) v4.0.0. Genes were classified as intolerant to loss of function (LoF) variants if LOEUF< 0.6 or pLI ≥0.9. We also used DGIdb v.5.0.1[32] to detect any druggable genes amongst our set of high confidence genes for BD risk.

QTL integrative analyses

Union consensus SNPs were investigated for putative causal relationships with BD via brain gene expression, splicing or methylation, using Summary data-based Mendelian randomization (SMR) (v1.03)[33,34]. Data on expression quantitative trait loci (eQTLs) and splicing quantitative trait loci (sQTLs) were obtained from the BrainMeta study (v2), which comprised RNA-seq data of 2,865 brain cortex samples from 2,443 unrelated individuals of EUR ancestries with genome-wide SNP data[35]. Data on methylation quantitative trait loci (mQTLs) were obtained from the Brain-mMeta study[36], a meta-analysis of adult cortex or fetal brain samples, comprising 1,160 individuals with methylation levels measured using the Illumina HumanMethylation450K array. We analyzed *cis*-QTLs, which were defined as those within 2 Mb of each gene[35]. Of the 22 union consensus SNPs, 10 were present in the BrainMeta QTL data, and 10 were present in the Brain-mMeta data. Using the BD GWAS[16] and QTL summary statistics[35], each union consensus SNP was analyzed as the target SNP for probes within a 2 Mb window on either side using the *--extract-target-snp-probe* option in SMR. Only probes for which the union consensus SNP was a genome-wide significant QTL ($P < 5 \times 10^{-8}$) were analyzed, to ensure robustly associated instruments for the SMR analysis[33,34]. A Bonferroni correction was applied for 579 probes tested in the eQTL ($P_{SMR} < 8.64 \times 10^{-5}$), 2,257 tests in the sQTL ($P_{SMR} < 2.21 \times 10^{-5}$) and 45 tests in the mQTL analyses ($P_{SMR} < 1.11 \times 10^{-3}$). The significance threshold for the HEIDI test (heterogeneity in dependent instruments) was $P_{HEIDI} \geq 0.01$[34]. The HEIDI test is used to identify potential violations of the Mendelian Randomization assumption, specifically the assumption of no horizontal pleiotropy. A SNP with passing the Bonferroni-corrected P_{SMR} and the P_{HEIDI} thresholds indicates either a direct causal role or a pleiotropic effect of the BD-associated SNPs on gene expression, splicing or methylation level.

Overlap with epigenomic peaks and rare variant association signal

Union consensus SNPs were examined for physical overlap with promoters or enhancers of gene expression in human brain cell-types. Data on epigenomic peaks were obtained from

purified bulk, H3K27ac and H3K4me3 ChIP-seq of neurons and astrocytes previously published and used to detect active promoters and enhancers[37]. Physical overlap was visually examined via locus plots using R (R version 4.1.2). For SNPs located in promoters, we assigned the corresponding gene name. For active enhancers, the target gene was assigned based on PLAC-Seq data[37] on enhancer-promoter interactions. Genes linked to union consensus SNPs via overlap with epigenomic peaks, SMR, or missense annotations, were further assessed for convergence with findings from an exome sequencing study of BD published by the Bipolar Exome (BipEx) Collaboration[38]. Using the BipEx browser[38], genes annotated to union consensus SNPs were compared for an overlap against BipEx genes characterized by a significant ($P < 0.05$) burden of either damaging missense or LoF variants.

Fine-mapping the MHC locus

The major histocompatibility complex (MHC) locus was fine-mapped separately due to its complex genetic variation and long-range LD structure[39]. The human leukocyte antigen (HLA) alleles and amino acid variants were imputed in the PGC BD data, using the 1000 Genomes phase 3 reference panel comprising 503 EUR individuals[40] with HLA alleles determined via sequencing. This reference was obtained from the CookHLA GitHub repository[41] (CookHLA v.1.0.1) and included 151 HLA alleles (65 2-digit and 86 4-digit) with a MAF >0.01 and <0.99, 1,213 amino acid variants, and 1,268 SNPs within the MHC region (chromosome 6, 29-34 Mb).

Variation in the MHC was imputed for 48 BD cohorts where individual-level genotyped SNP data were available within the PGC (33,827 BD, 53,953 controls), using IMPUTE2, implemented via the Rapid Imputation and COmputational PIpeLIne for GWAS (RICOPILI)[42]. RICOPILI was used to perform association analysis, under an additive logistic regression model in PLINK v1.90[28], covarying for the first 5 principal components (PCs) of genetic ancestry and any others associated with case-control status within each cohort, as per the BD GWAS[16]. To control test statistic inflation at variants with low MAF in small cohorts, variants were retained only if cohort MAF was greater than 1% and minor allele count was greater than 10 in either cases or controls (whichever had smaller N). Meta-analysis of the filtered association statistics was conducted using an inverse-variance-weighted fixed-effects model in METAL (version 2011-03-25) via RICOPILI[43].

Conditional analysis of the MHC-association results was performed to identify whether there are any additional independent associations, by conditioning on the top lead variant within the locus. In brief, the dosage data for the top lead variant in the meta-analysis were extracted for each cohort, converted into a single value representing the dosage of the A1 allele (range 0-2), and this was added as a covariate in the analysis. Association testing, filtering of results per cohort, and the meta-analysis were carried out as described above.

Polygenic risk scoring

Fine-mapping results were further evaluated by testing whether fine-mapping effect sizes could improve the performance of PRS in independent cohorts using PolyPred[44], a method which combines effect sizes from fine-mapping with those from a standard PRS approach, such as PRS-CS[45]. PRS were calculated for individuals in 12 testing cohorts of BD cases and controls that were independent of the BD GWAS: three new PGC cohorts of EUR ancestries, two cohorts of East Asian ancestries, four cohorts of admixed-African American ancestries, and three cohorts of Latino ancestries, some of which have been described previously[16] (**Supplementary Note**).

An analytical workflow outlining the steps of the PolyPred pipeline that we followed is shown in **Supplementary Figure S1**. First, the standard approach used was PRS-CS, which uses a Bayesian regression framework to place continuous shrinkage priors on effect sizes of SNPs in the PRS, adaptive to the strength of their association signal in the BD GWAS[16], and the LD structure from an external reference panel[45]. The UKB EUR ancestry reference panel was used to estimate LD between SNPs, matching the ancestry of the discovery GWAS[16]. PRS-CS yielded weights for approximately 1 million SNPs to be included in the PRS. Second, genome-wide fine-mapping was performed on the BD GWAS summary statistics[16], using both SuSiE and PolyFun-SuSie as previously described with LD information obtained from the HRC reference panel, to derive causal effect sizes for all SNPs across the genome. Third, PolyPred was used to combine the SNP weights from PRS-CS with SuSie effect sizes ("SuSie+PRS-CS") and SNP weights from PRS-CS with PolyFun-SuSiE effect sizes ("Polypred-P"). Briefly, Polypred "mixes" the effect sizes from the two predictors via the non-negative least squares method, assigning a weight to each predictor that yields the optimally performing PRS in a specific testing cohort. Each testing cohort was used to tune the optimal PolyPred weights. Fourth, three PRS were calculated for each individual in the testing cohorts, using PLINK v1.90[28] to weight SNPs by their effect sizes from PRS-CS, SuSiE+PRS-CS and Polypred-P respectively, and sum across all SNPs in each PRS. Finally, PRS were tested for association with case versus control status in each testing cohort using a logistic regression model including PCs 1-5 and any others as necessary to control for genetic ancestry[46]. In each testing cohort, the amount of phenotypic variance explained by the PRS (R^2) and the 95% confidence intervals were calculated on the liability scale[47], using the r2redux R package[48], assuming a lifetime prevalence of BD in the general population of 2%. The R^2 of each fine-mapping-informed PRS was statistically compared against the R^2 of PRS-CS using the r2redux package (r2_diff function)[48].

Results

Fine-mapping identifies likely causal BD variants

Stepwise conditional analyses using GCTA-COJO were performed in each of the 64 BD GWS loci (**Supplementary Table S2**), conditioning associations on their top lead SNP and any subsequent conditionally independent associations, to identify loci that contained independent signals (conditional $P < 5 \times 10^{-6}$). This analysis supported the existence of one association signal at 62 loci (**Supplementary Table S3**), and two independent association signals within the MSRA locus on chromosome 8 and the RP1-84O15.2 locus on chromosome 8 (**Supplementary Table S3**).

Excluding the MHC, GWS loci were fine-mapped via a suite of Bayesian fine-mapping tools (SuSiE, FINEMAP, PolyFun+SuSiE, PolyFun+FINEMAP) to prioritize SNPs likely to be causal for BD, and examine the impact of different LD reference panels and fine-mapping windows (see Methods). **Table 1** shows the number of SNPs with a PIP >0.95 and PIP >0.50 in each fine-mapping analysis. Generally, incorporating functional priors using PolyFun yielded greater numbers of SNPs passing these thresholds (mean 52% increase). Typically, more SNPs exceeded these PIP thresholds when using the larger 3 Mb fine-mapping window versus the GWS locus window (mean 25% increase). As anticipated, making the simple assumption of a single-causal variant per locus to comply with the conditional analysis and not leveraging LD information prioritized the fewest SNPs. Incorporating an LD reference panel to allow for multiple causal SNPs within a locus prioritized more SNPs, with the UKB reference panel yielding the most SNPs with PIPs exceeding our specified thresholds.

Approximately a third of GWS loci (N= 20) had high PIP SNPs (>0.50). Employing different fine-mapping methods and LD reference panels revealed a substantial number of consensus SNPs with PIP >0.50, particularly with a 3 Mb window, but fewer met the stricter threshold of PIP >0.95. The number of 95% credible sets per locus varied based on the fine-mapping method (**Supplemental Note, Supplemental Figure S2, Supplemental Figure S3)**.

The union consensus set (PIP >0.5) comprised 22 SNPs (from 20 GWS loci), indicating that many of the same SNPs were prioritized regardless of which LD reference panel or fine-mapping window was used (**Figure 2**). There were 9 SNPs consistently prioritized as the likely causal variant across all LD reference panels and fine-mapping windows (**Figure 2, Supplementary Figure S3**). The distribution of SNPs with PIP >0.95 for each GWS locus across different methods, LD reference panels and fine-mapping windows is provided in **Supplemental Table S4**.

Fine-mapping method:	SuSiE	FINEMAP	PolyFun+SuSiE	PolyFun+FINEMAP	Consensus SNPs
	UKB LD - 3 Mb window				
N SNPs with PIP > 0.95	4	6	6	7	6
N SNPs with PIP > 0.50	8	18	14	29	20
	UKB LD - GWS locus window				
N SNPs with PIP > 0.95	2	7	5	9	8
N SNPs with PIP > 0.50	7	17	13	22	20
	HRC LD - 3 Mb window				
N SNPs with PIP > 0.95	4	5	6	5	6
N SNPs with PIP > 0.50	8	16	14	22	17
	HRC LD - GWS locus window				
N SNPs with PIP > 0.95	2	2	4	2	2
N SNPs with PIP > 0.50	7	9	11	15	14
	In-sample LD - 3 Mb window				
N SNPs with PIP > 0.95	4	6	6	5	6
N SNPs with PIP > 0.50	8	16	13	22	16
	In-sample LD - GWS locus window				
N SNPs with PIP > 0.95	3	2	4	4	2
N SNPs with PIP > 0.50	8	9	10	18	15
	No LD - 3 Mb window				
N SNPs with PIP > 0.95	2	2	2	2	2
N SNPs with PIP > 0.50	7	8	16	16	17
	No LD - GWS locus window				
N SNPs with PIP > 0.95	2	2	3	3	3
N SNPs with PIP > 0.50	8	8	15	15	16

Table 1. Number of SNPs with posterior inclusion probability (PIP) >0.95 or >0.50 in each fine-mapping analysis. Results are shown for each of the four fine-mapping methods, the 3 linkage disequilibrium (LD) reference panels, single-variant fine-mapping with no LD, and the two fine-mapping windows (3 Mb window and the genome-wide significant (GWS) locus window). Within a cell, all SNPs are from different GWS loci. Consensus SNPs were defined as those with PIP >0.95 or PIP >0.50 and in the 95% credible set in 2+ fine-mapping methods, when using the same LD reference panel and fine-mapping window (except for the single-variant fine-mapping analysis, since the 95% credible set criterion does not apply). The union consensus set (defined as consensus SNPs with PIP >0.50), comprised 22 SNPs from 20 GWS loci. HRC - Haplotype Reference Consortium, UKB - UK Biobank.

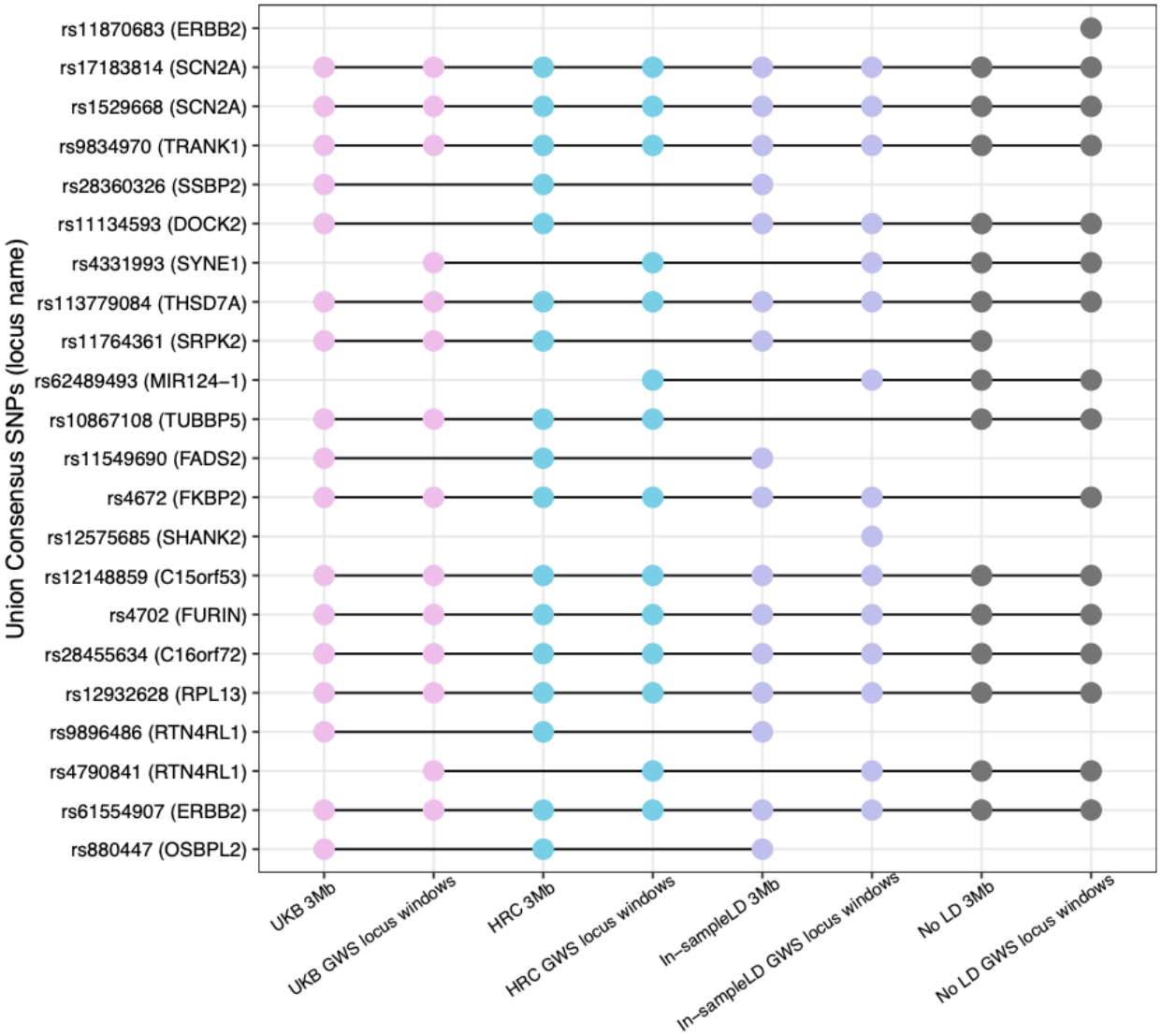


Figure 2. Convergent evidence across fine-mapping approaches. The y-axis shows the 22 SNPs in the union consensus set and the name of each genome-wide significant locus (as defined in the original GWAS). The x-axis shows the linkage disequilibrium (LD) reference panel (color-coded; UK Biobank (UKB) - pink, Haplotype Reference Consortium (HRC) - blue, in-sample LD - purple, and no LD - grey) and the fine-mapping window. Each point represents a consensus SNP with PIP >0.50 in 2+ fine-mapping methods, within the corresponding approach.

Characterization of union consensus SNPs

Union consensus SNPs were initially characterized using standard metrics from online databases. Variant annotation of the union consensus SNPs via VEP[30] indicated that 7 of the 22 fall in intronic regions (**Supplementary Table S5**). There were 20 SNPs annotated to genes using VEP, and 16 of which were the closest gene to the index SNP (**Supplementary Table S5**). Three of the union consensus SNPs are missense variants: rs17183814 in *SCN2A* (CADD: 20, ClinVar

benign), rs4672 in *FKBP2* (CADD: 22.5, not in ClinVar) and rs11549690 in *TRPT1* (CADD: 19.8, not in ClinVar). According to RegulomeDB annotations, two union consensus SNPs likely have regulatory functions: rs4331993 (intronic in *SYNE1*) and rs10867108 (3' untranslated region of *CACNA1B*) (**Supplementary Table S5**). Of the protein coding genes prioritized by VEP annotations, 10 of the 18 had high pLI scores (≥0.9) and 12 had low LOEUF scores (<0.65), indicating intolerance to LoF variants (**Figure 3**).

QTL integrative analyses and overlap with epigenomic peaks

Summary data-based Mendelian randomization (SMR)[33,34] was used to identify putative causal relationships between union consensus SNPs and BD via gene expression, splicing or methylation, by integrating the BD GWAS association statistics with brain eQTL, sQTL and mQTL summary statistics. eQTL and sQTL data were based on the BrainMeta study (2,865 brain cortex samples from 2,443 unrelated individuals of EUR ancestries)[35] and mQTL data were from the Brain-mMeta study (adult cortex or fetal brain samples in 1,160 individuals)[36]. Union consensus SNPs with genome-wide significant cis-QTL P values ($P < 5x10^{-8}$) and their corresponding gene expression, slicing or methylation probes were selected as target SNP-probe pairs for SMR, yielding 579, 2,257 and 45 SNP-probe pairs for eQTL, sQTL and mQTL analyses respectively. In the eQTL analyses, there were 6 union consensus SNPs with significant P_{SMR} that passed the HEIDI (heterogeneity in dependent instruments) test for 10 different genes, suggesting that their effect on BD is mediated via gene expression in the brain (**Figure 3, Supplementary Table S6**). Four of the union consensus SNPs showed evidence of causal effects on BD via expression of more than one gene in their *cis*-region. In the sQTL analyses, there were 8 union consensus SNPs with significant P_{SMR} results, and passing the HEIDI test, implicating 12 genes (**Figure 3, Supplementary Table S6**). In the mQTL analyses, there were 24 SNP-probe pairs passing the P_{SMR} and P_{HEIDI} thresholds; of which two methylation probes were annotated to specific genes (*FKBP2* and *PLCB3*) (**Figure 3, Supplementary Table S6**).

There were 12 union consensus SNPs that physically overlapped with active enhancers or promoters of gene expression in brain cell-types[37], particularly neurons (**Figure 3**). Four union consensus SNPs were located in active promoters of the *SCN2A, THSD7A, FKBP2* and *THRA* genes. Through the utilization of PLAC-seq data, we explored enhancer-promoter interactions, specifically for enhancers in which there is a physical overlap with the union consensus SNPs, and prioritized their genes (**Figure 3**). Amongst the implicated target genes through enhancer-promoter interactions are *INSYN2B*, *SYNE1*, *RASGRP1*, *CRTC3*, *DPH1* and *KCNQ2*.

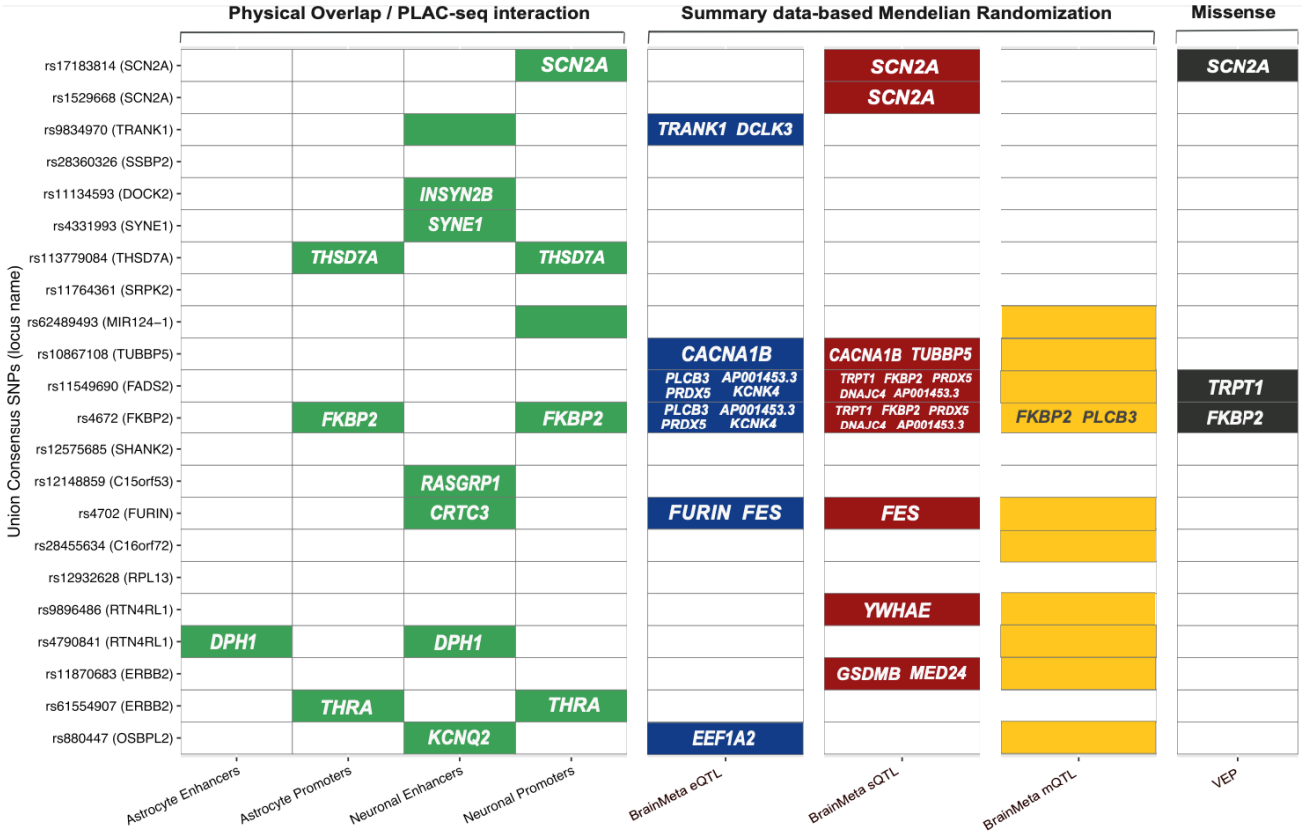


Figure 3. Summary of analyses performed to link each fine-mapped SNP to the relevant gene(s). The y-axis shows the 22 union consensus SNPs and the name of the corresponding genome-wide significant locus (as defined in the original GWAS). On the x-axis, the columns depict the results of 8 analyses performed to link the fine-mapped SNPs to the relevant gene(s). The analysis method and the dataset used are labeled above and below the figure respectively. Colored cells denote significant results and the relevant gene names are printed within each cell. For fine-mapped SNPs located in active enhancers, the relevant genes were obtained using data on PLAC-seq interactions with gene promoters. A colored cell includes no gene name when there was no known interaction between the enhancer and a promoter, or when the methylation probe was not annotated to any gene. Empty cells are those with non-significant results, or where the SNP was not present in the dataset used.

Candidate risk genes based on convergence of evidence

By aggregating multiple lines of fine-mapping validation evidence, we present results for high-confidence genes for BD. Specifically, a gene was characterized as high-confidence if it was linked to a fine-mapped SNP via active promoters or enhancers, brain gene expression, splicing or methylation, or if the fine-mapped SNP was a missense variant (**Figure 3**, **Supplementary Figure S5**). Taken together, the data support the roles of the following 25 genes in BD: *SCN2A, TRANK1, DCLK3, INSYN2B, SYNE1, THSD7A, CACNA1B, TUBBP5, PLCB3, PRDX5, KCNK4, AP001453.3, TRPT1, FKBP2, DNAJC4, RASGRP1, FURIN, FES, YWHAE, DPH1, GSDMB, MED24,*

THRA, EEF1A2, and *KCNQ2*. **Supplementary Figure S5** provides multi-track locus plots depicting GWAS association statistics, fine-mapping results, overlap with epigenomic peaks from neurons or astrocytes and gene tracks for 11 loci as examples. We assessed the high-confidence genes for evidence of rare variant associations with BD, using data from the BipEx exome sequencing study[38]. Amongst the 25 genes examined, *THSD7A*, *CACNA1B*, *SCN2A* and *TRANK1* had a significant burden ($p < 0.05$) of damaging missense or LoF variants in BD versus controls. Many high-confidence genes were classified as druggable based on DGIdb v5.0.3 (*SCN2A*, *CACNA1B, PRDX5, THRA, EEF1A2, KCNQ2* and *FES*).

Dissecting the MHC locus

We performed association analyses of variants in the MHC region (chromosome 6, 29-34 Mb) including *HLA* alleles, amino acids, SNPs and insertion/ deletion variants, in a sample of 33,781 BD cases and 53,869 controls. The most significant variant in the MHC was rs1541269, located at 30.1 Mb (OR A allele = 1.12, 95% CI = 1.08-1.15, P = 6.71×10^{-12}). This SNP is in moderate LD with the top lead MHC SNP in the full BD GWAS (rs13195402, 26.5 Mb, LDlink EUR $r^2 = 0.55$)[16], which was not present in the imputation reference panel used here. After conditioning on the top lead variant in the MHC from this association analysis (rs1541269), no variants in the MHC remained GWS, suggesting a single association signal across the region in BD (**Supplementary Figure S4B, Supplementary Table S8**). Prior to conditioning, there were 10 variants in *HLA* genes reaching GWS ($P < 5 \times 10^{-8}$) , including 3 classical *HLA* alleles, 2 amino acid variants and 5 intragenic *HLA* SNPs (**Supplementary Figure S4A, Supplementary Table S7)**. *HLA-C*0701* showed a protective effect on BD (OR = 0.91, 95% CI = 0.88-0.94, P = 3.40×10^{-10}) as did *HLA-B*0801* (OR = 0.91, P= 4.08×10^{-8}) (**Supplementary Figure S4A, Supplementary Table S7**). An amino acid change from lysine to asparagine at position 66 in *HLA-C* and the presence of aspartic acid at position 9 in *HLA-A* also had protective effects on BD and were GWS (**Supplementary Table S7).** However none of the HLA variants remained GWS after conditioning on the top lead SNP, rs1541269.

Leveraging fine-mapping to improve polygenic risk scoring

We assessed whether fine-mapping results could be used to improve the performance of BD PRS in 12 testing cohorts: three EUR cohorts that were independent of the BD GWAS, two East Asian cohorts, four admixed African American cohorts, and three Latino cohorts[46,49,50]. Standard PRS were calculated using the PRS-CS method, and fine-mapping informed PRS were calculated via PolyPred, to integrate statistical fine-mapping results (SuSiE+PRS-CS) or functional fine-mapping results (Polypred-P). Across PRS methods, PRS were significantly higher in BD cases versus controls in all EUR target cohorts and most non-EUR cohorts (**Figure 4, Supplementary Table S9**). Using PRS-CS, the mean phenotypic variance explained on the liability scale was 11.42% in EUR ancestries, 2.39% in East Asian ancestries, 0.14% in African American ancestries

and 0.30% in Latino ancestries (**Figure 4**, **Supplementary Table S9)**. Examining fine-mapping-informed PRS, SuSiE+PRS-CS or Polypred-P explained more phenotypic variance than PRS-CS in 9/12 cohorts, with PolyPred-P typically showing the best performance (**Figure 4**). However, increased variance explained by SuSiE+PRS-CS or Polypred-P compared with PRS-CS, was only statistically significant in the Japanese BD cohort (P = $1.22x10^{-5}$ and P = $2.29x10^{-6}$ respectively), one African American (P = 0.035 and P = 0.044 respectively) and one Latino cohort (P = 0.046 and P = 0.002 respectively) (**Supplementary Table S9, Figure 4).**

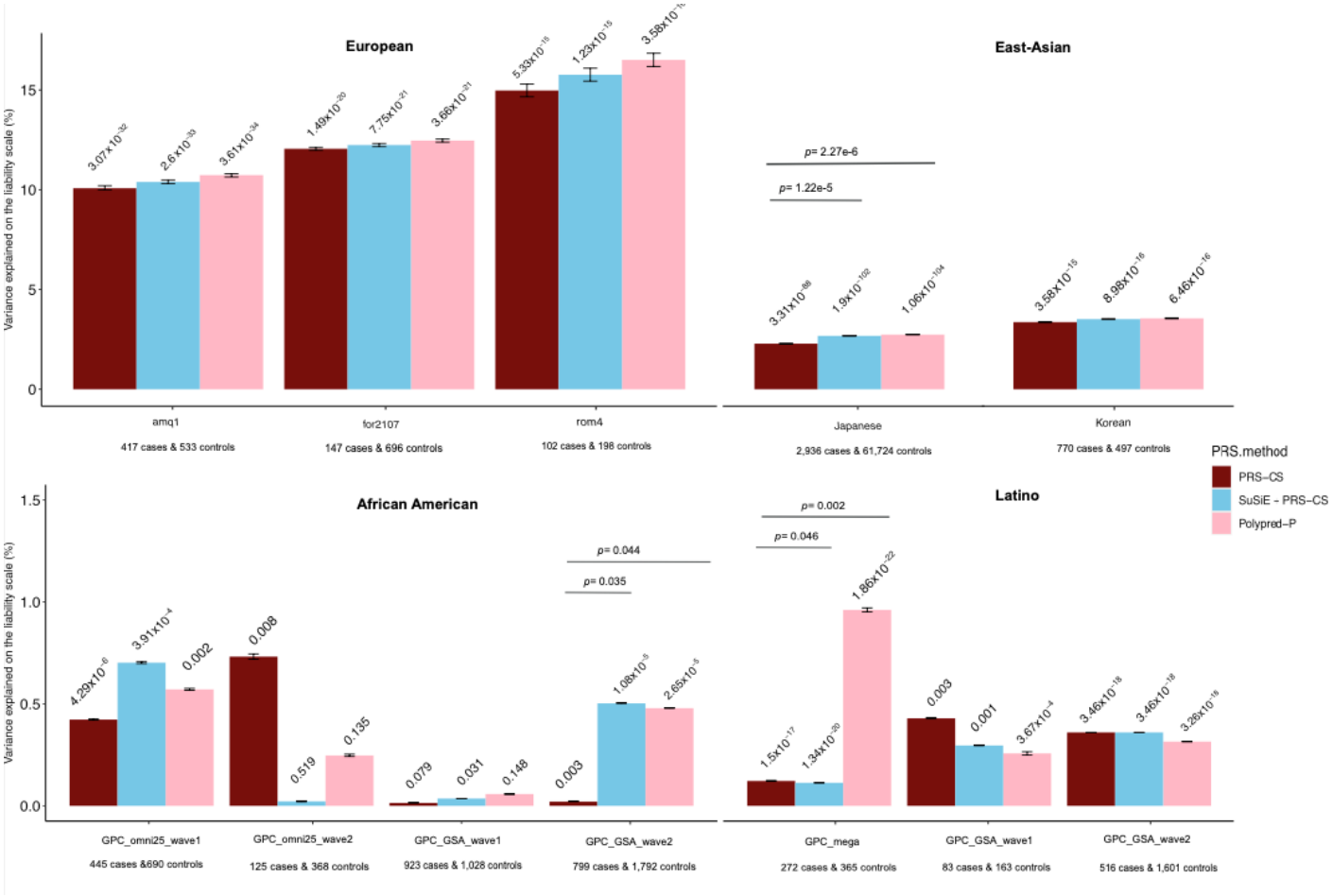


Figure 4. **Phenotypic variance in BD explained by standard PRS (PRS-CS) and fine-mapping-informed PRS (SuSiE+PRS-CS and PolyPred-P) in target cohorts of diverse genetic ancestries.** The x-axis displays the target cohorts, grouped by genetic ancestry, and the PRS method used. The name of each cohort and the number of BD cases and controls is shown below each barplot. The y-axis shows the percentage variance explained on the liability scale, assuming a 2% population prevalence of BD. Error bars represent 95% confidence intervals on the variance explained. P-values for the association of PRS with case versus control status are printed on top of each bar. Significant P-values (P < 0.05) for the test of difference in variance explained by the fine-mapping informed PRS versus PRS-CS are provided above the horizontal lines.

Discussion

In the most comprehensive fine-mapping study of BD GWAS risk loci to date, we applied a suite of statistical and functional fine-mapping methods to prioritize 22 likely causal SNPs for BD in 20 genomic loci. We linked these SNPs to genes and investigated their likely functional consequences, by integrating variant annotations, brain cell-type epigenomic annotations, brain QTLs, and results from exome sequencing in BD. Convergence of evidence across these analyses prioritized 25 high-confidence genes, which are strong candidates for functional validation experiments to understand the mechanisms by which they increase risk of BD.

We defined a union consensus set of SNPs representing those likely causal for BD based on the convergence between fine-mapping methods, LD reference panels and fine-mapping windows. This comprised 22 SNPs (from 20 GWS loci), indicating that many of the same SNPs were prioritized across fine-mapping analyses (**Figure 2**). Linking these SNPs to genes and investigating their likely functional consequences using computational approaches and relevant datasets, prioritized 25 high-confidence genes (**Figure 3**). The *SCN2A* (Sodium channel protein type 2 subunit alpha) gene had two union consensus SNPs, one intronic and another a missense variant located in a neuronal promoter. SCN2A has been implicated in epilepsy and neurodevelopmental disorders[51]. *THSD7A* encodes N-glycoprotein thrombospondin type 1 domain containing 7A, which mediates endothelial cell migration, tube formation and in neuroangiogenesis[52]. The gene is highly expressed in inhibitory and excitatory neurons and has been previously associated with treatment response to antiepileptic drug mood stabilizers amongst BD patients[53]. *CACNA1B* (Calcium Voltage-Gated Channel Subunit Alpha1 B) was prioritized through fine-mapping of rs10867108 which is located in the 3' UTR of the gene and SMR analysis suggested this SNP increases risk of BD through increased cortical expression of *CACNA1B*. *SCN2A*, *THSD7A* and *CACNA1B* were also all supported by the BipEx exome sequencing study, with a significant burden ($p < 0.05$) of damaging missense or LoF variants in BD versus controls[38]. Our findings align with the results of drug target enrichment analyses performed on the BD GWAS results, which found significant enrichment of associations amongst genes encoding targets of calcium channel blockers and antiepileptics[16]. Taken together, these genes and pathways warrant functional investigation to understand biological mechanisms and potential for therapeutic modulation.

The *FURIN* and *TRANK1* genes had strong evidence for a causal role in BD, and both have been studied previously in relevant functional experiments. Here, *FURIN* was prioritized through fine-mapping of rs4702, which overlaps a neuronal enhancer, with SMR suggesting that the BD risk allele acts via decreasing cortical expression of FURIN. This mechanism has been studied via CRISPR editing in human induced pluripotent stem cell (iPSC)-derived neurons, where allelic

conversion to the BD risk allele in excitatory neurons resulted in decreased FURIN expression, neurite length, and firing duration[54]. In our study, the nearby *FES* gene was also prioritized, as SMR indicated that rs4702 plays a role in cortical expression and alternative splicing of FES. This gene encodes a tyrosine kinase which is involved in many aspects of cellular differentiation. The prioritization of two genes in the same locus, underscores that one fine-mapped SNP may have multiple functional consequences. *TRANK1* was implicated as a high confidence gene through the role of the fine-mapped SNP (rs9834970) in cortical expression, and the burden of damaging rare variants identified in BipEx. iPSC-derived neural progenitor cells carrying the BD risk allele at this SNP, have been shown to exhibit lower *TRANK1* expression than homozygotes for the non-risk allele[55]. Furthermore, this experiment showed that *TRANK1* expression levels could be rescued by chronic treatment with therapeutic doses of valproic acid, an antiepileptic which is also used to treat mania. Other high-confidence genes include *FKBP2*, which plays an important role in immunoregulation, protein folding and trafficking[56] and *SYNE1*, with one of its transcripts encoding CPG2, a brain-specific protein localized to excitatory postsynaptic sites, which regulates glutamate receptor internalization[57]. PLAC-seq data also prioritized genes with relevant biological functions, such as *INSYN2B* (inhibitory synaptic factor family member 2b), *KCNQ2* (potassium voltage-gated channel subfamily Q member 2) and *DPH1* (diphthamide biosynthesis 1) with variants in this gene associated with developmental delay[58]. Of the 25 high-confidence genes, *CACNA1B, TRPT1*, *YWHAE* and *EEF1A2* were not the closest gene to the GWAS index SNP, and of the 22 union consensus SNPs, 8 were not the top lead SNP from the GWAS, illustrating the utility of fine-mapping. We observed that only two genes (*FURIN*, *DCLK3*) overlapped with genes associated with schizophrenia. This could be explained by the polygenic nature of both disorders and the limited and partial overlap (N = 15) of the 64 BD GWS loci against the 287 SCZ GWS loci[23].

In the MHC, there were several polymorphic alleles and amino acid variants in the *HLA-C* and *HLA-B* genes associated with BD at GWS (chromosome 6, 31.2-31.3 Mb). The *HLA-C*07:01* and *HLA-B*08:01* alleles were negatively associated with BD, in line with previous studies reporting their protective effects on SCZ[59,60]. However, these associations were removed after conditioning on the top lead variant in the MHC (rs1541269, 30 Mb), suggesting the effects were driven by LD with more strongly associated variants located upstream. This is consistent with published findings in the PGC BD data, showing no association between the structural variants in the complement component 4 genes (C4A/C4B) (~31.9 Mb) and BD, either before or after conditioning on the most associated MHC SNP (rs13195402, 26.4 Mb)[16]. Overall, this analysis of HLA variation in BD again suggests a single association signal across the MHC, and that the causal variants and genes are outside the classical MHC locus, in contrast to findings in schizophrenia[61].

Fine-mapping-informed PRS, developed by combining GWAS effect sizes and genome-wide fine-mapping effect sizes using PolyPred, generally explained a greater proportion of phenotypic variance compared with PRS based on GWAS effect sizes alone. This was true in all EUR and East Asian ancestry testing cohorts, and some admixed African American and Latino cohorts (**Figure 4, Supplementary Table S9**). The increase in phenotypic variance explained by fine-mapping-informed PRS adds support to our fine-mapping results, as leveraging information on causal effect sizes rather than relying solely on association statistics is expected to improve genetic risk prediction. However, the increase in the phenotypic variance explained by fine-mapping-informed PRS versus PRS-CS was only statistically significant in 3 of 12 tested cohorts. For some cohorts, this may be explained by their modest sample sizes, but in one admixed African American cohort and three Latino cohorts, fine-mapping-informed PRS explained less phenotypic variance than PRS-CS. These results suggest that PolyPred does not perform well in admixed ancestries. Despite incorporating causal effect sizes from genome-wide fine-mapping, with the expectation that causal variants are shared across ancestries, the performance of all PRS in non-European cohorts still lagged greatly behind that in Europeans. Nevertheless, we anticipate that the sharing of genome-wide fine-mapping results in a manner analogous to GWAS summary statistics will facilitate calculation of fine-mapping-informed PRS in many cohorts, and the development of improved PRS methods, particularly for admixed populations.

Our strategy of applying a suite of fine-mapping methods and examining the convergence of the results was driven by the variety of the underlying fine-mapping algorithms, and their corresponding strengths and limitations. Consistent with previous literature, we detected more SNPs with high PIPs when incorporating functional priors using PolyFun[18]. As expected, the specification of LD structure, fine-mapping window, and number of causal variants impacted fine-mapping results. Considering "in-sample" LD from the PGC BD data (albeit a subset of cohorts that were available) as the gold-standard, using the HRC reference panel yielded the most similar fine-mapping results. This observation may be explained by the HRC being used as an imputation reference panel for almost all cohorts in the GWAS. Results suggest that a large and well-matched LD reference panel to the GWAS sample can be used to achieve high-quality fine-mapping results. This has advantageous implications in scenarios when calculating in-sample LD is not possible due to data sharing restrictions, or when obtaining LD information from many cohorts becomes increasingly challenging as GWAS meta-analyses grow. Moreover, although conditional analysis indicated 1 causal variant per GWS locus, our results are highly consistent when using LD reference panels and allowing up to 5 causal variants per GWS locus. The latter analyses also yielded more likely causal SNPs. To facilitate rapid and scalable fine-mapping of GWAS loci, we developed a fine-mapping pipeline (GitHub: https://github.com/mkoromina/SAFFARI) with options to specify multiple fine-mapping methods, GWAS summary statistics, fine-mapping windows, and LD reference panels.

Several limitations of this study and future directions must be noted. First, our fine-mapping focused exclusively on EUR ancestry data, owing to the composition of the PGC BD GWAS. However, this enabled us to investigate the impact of LD reference panels on fine-mapping, which would be challenging for diverse ancestry data, given the limited availability of such panels at present. Increasing ancestral diversity in BD GWAS is an active area of research[46], and in future the differences in LD structure between populations could be leveraged to aid fine-mapping[62] and PRS predictions[44]. Second, we approximated "in-sample LD" of the GWAS as we only had access to a subset of the individual-level data (73% of the total effective sample size), we used best guess genotypes to represent imputed dosages, and we merged genotypes across cohorts and calculated LD, in contrast to the GWAS, which was a meta-analysis between cohorts. Third, this study prioritized likely causal variants or genes at 20 of the 64 GWS loci, meaning that many loci were not robustly fine-mapped. Our approach was conservative in that we focused on SNPs with high PIPs (>0.50), that were part of credible sets, and were supported by different fine-mapping models. The improvements in PRS performance after integrating genome-wide fine-mapping results, suggest that our analyses capture meaningful information on causality in other genomic regions that did not meet the stringent criteria we applied to fine-map GWS loci. Fourth, these statistical analyses prioritize variants and genes with high-probabilities of being causal risk factors for BD, however computational approaches fall short of proving causality, and have limited capacity to uncover mechanisms. Finally, the enhancer, promoter and QTL data used may be incomplete due to cell-type or context-specific effects, or incomplete mapping of active enhancers to their target genes, and therefore some union consensus SNP effects may not have been detected in our analysis.

In summary, we conducted a comprehensive statistical and functional fine-mapping analysis of BD genomic loci, yielding a resource of likely causal genes and variants for the disorder. These genes and variants now require investigation in functional laboratory experiments to validate their roles, understand mechanisms of risk, and examine opportunities for therapeutic intervention in BD.

Data availability

The PGC's policy is to make genome-wide summary results public. Genome-wide fine-mapping results will be made available through the PGC website upon publication (https://www.med.unc.edu/pgc/results-and-downloads). Individual-level genetic data are accessible via Secondary 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.

Code availability

Analysis scripts are available online at [Github: https://github.com/mkoromina/SAFFARI]. All software used is publicly available at the URLs or references cited.

Acknowledgements

For the purposes of open access, the author has applied a Creative Commons Attribution (CC BY) license to any Accepted Author Manuscript version arising from this submission. We thank the participants who donated their time, life experiences and DNA to this research and the clinical and scientific teams that worked with them. This project was funded by the Baszucki Brain Research Fund via the Milken Institute Center for Strategic Philanthropy. We are deeply indebted to the investigators who comprise the PGC. The PGC has received major funding from the US National Institute of Mental Health (PGC4: R01MH124839, PGC3: U01 MH109528; PGC2: U01 MH094421; PGC1: U01 MH085520). Statistical analyses were carried out on the NL Genetic Cluster Computer (http://www.geneticcluster.org) hosted by SURFsara and the Mount Sinai high performance computing cluster (http://hpc.mssm.edu), which is supported by the Office of Research Infrastructure of the National Institutes of Health under award numbers S10OD018522 and S10OD026880. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Full acknowledgements are included in the Supplementary Note. JRIC is supported by a grant from the Medical Research Foundation (MRF-001-0012-RG-COLE-C0930) and by the NIHR Maudsley Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the authors and not necessarily those of the NIHR or the UK Department of Health and Social Care. EV thanks the support by CIBER - Consorcio Centro de Investigación Biomédica en Red- (CB07/09/0004), Instituto de Salud Carlos III, Spanish Ministry of Science and Innovation and grants PI18/00805 and PI21/00787, integrated into the Plan Nacional de I+D+I and co-financed by the ISCIII-Subdirección General de Evaluación and the Fondo Europeo de Desarrollo Regional (FEDER); the Instituto de Salud Carlos III; the Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement (2021 SGR 01358), the CERCA Programme, and the Departament de Salut de la Generalitat de Catalunya for the PERIS grant SLT006/17/00357. Thanks also for the support of the European Union Horizon 2020 research and innovation program (EU.3.1.1. Understanding health, wellbeing and disease: Grant No 754907 and EU.3.1.3. Treating and managing disease: Grant No 945151). PBM has been funded through an Australian National Health and Medical Research Council Investigator Grant (1177991). Work for the Japanese cohort was supported by Japan Agency for Medical Research and Development (AMED) grants 22wm0425008, 21ek0109555,

21tm0424220, 21ck0106642, 23ek0410114 and 23tm0424225, Japan Society for the Promotion of Science (JSPS) KAKENHI grant 21H02854 and JP20H00462. Work for the 'for 2107' cohort was supported through funds by the German Research Foundation (DFG grants FOR2107 KI588/14-1, and KI588/14-2, and KI588/20-1, KI588/22-1 to Tilo Kircher, Marburg, Germany). Biosamples and corresponding data were sampled, processed and stored in the Marburg Biobank CBBMR. Ethics approval was obtained from the ethics committees of the Medical Schools of the Universities of Marburg (approval identifier Studie 07/2014) and Münster, respectively, in accordance with the Declaration of Helsinki. All subjects volunteered to participate in the study and provided written informed consent.

Competing interests

OAA has served as a speaker for Janssen, Lundbeck, and Sunovion and as a consultant for Cortechs.ai. SKS has served as speaker for Janssen, Takeda and Medice Arzneimittel Puetter GmbH & CoKG. EV has received grants and served as consultant, advisor or CME speaker for the following entities (unrelated to the present work): AB-Biotics, Abbott, Abbvie, Adamed, Angelini, Biogen, Biohaven, Boehringer Ingelheim, Casen-Recordati, Celon, Compass, Dainippon Sumitomo Pharma, Ethypharm, Ferrer, Gedeon Richter, GH Research, Glaxo Smith-Kline, Idorsia, Janssen, Johnson & Johnson, Lundbeck, Newron, Novartis, Organon, Otsuka, Rovi, Sage, Sanofi-Aventis, Sunovion, Takeda, and Viatris. PBM has received remuneration from Janssen (Australia) and Sanofi (Hangzhou) for lectures, and Janssen (Australia) for advisory board membership. MOD and MJO have received grants from Akrivia Health and Takeda Pharmaceuticals for work unrelated to this project.

References

1. O'Connell, K. S. & Coombes, B. J. Genetic contributions to bipolar disorder: current status and future directions. *Psychol. Med.* **51**, 2156–2167 (2021).
2. Craddock, N. & Sklar, P. Genetics of bipolar disorder. *Lancet* **381**, 1654–1662 (2013).
3. McGuffin, P. *et al.* The heritability of bipolar affective disorder and the genetic relationship to unipolar depression. *Arch. Gen. Psychiatry* **60**, 497–502 (2003).
4. Smoller, J. W. & Finn, C. T. Family, twin, and adoption studies of bipolar disorder. *Am. J.*

Med. Genet. C Semin. Med. Genet. **123C**, 48–58 (2003).

5. Stahl, E. A. *et al.* Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat. Genet.* **51**, 793–803 (2019).
6. 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).
7. Charney, A. W. *et al.* Evidence for genetic heterogeneity between clinical subtypes of bipolar disorder. *Transl. Psychiatry* **7**, e993 (2017).
8. Cichon, 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).
9. 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).
10. Green, E. K. *et al.* Association at SYNE1 in both bipolar disorder and recurrent major depression. *Mol. Psychiatry* **18**, 614–617 (2013).
11. 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).
12. Mühleisen, T. W. *et al.* Genome-wide association study reveals two new risk loci for bipolar disorder. *Nat. Commun.* **5**, 3339 (2014).
13. Scott, L. J. *et al.* Genome-wide association and meta-analysis of bipolar disorder in individuals of European ancestry. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 7501–7506 (2009).
14. 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).
15. Smith, E. N. *et al.* Genome-wide association study of bipolar disorder in European American and African American individuals. *Mol. Psychiatry* **14**, 755–763 (2009).
16. Mullins, N. *et al.* Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nat. Genet.* **53**, 817–829 (2021).

17. Schaid, D. J., Chen, W. & Larson, N. B. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat. Rev. Genet.* **19**, 491–504 (2018).
18. Weissbrod, O. *et al.* Functionally informed fine-mapping and polygenic localization of complex trait heritability. *Nat. Genet.* **52**, 1355–1363 (2020).
19. Schilder, B. M., Humphrey, J. & Raj, T. echolocatoR: an automated end-to-end statistical and functional genomic fine-mapping pipeline. *Bioinformatics* **38**, 536–539 (2022).
20. Wang, G., Sarkar, A., Carbonetto, P. & Stephens, M. A simple new approach to variable selection in regression, with application to genetic fine mapping. *J. R. Stat. Soc. Series B Stat. Methodol.* **82**, 1273–1300 (2020).
21. Benner, C. *et al.* FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* **32**, 1493–1501 (2016).
22. Kanai, M. *et al.* Insights from complex trait fine-mapping across diverse populations. *medRxiv* 2021.09.03.21262975 (2021) doi:10.1101/2021.09.03.21262975.
23. Trubetskoy, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**, 502–508 (2022).
24. Mölder, F. *et al.* Sustainable data analysis with Snakemake. *F1000Res.* **10**, 33 (2021).
25. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
26. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
27. Gazal, S. *et al.* Functional architecture of low-frequency variants highlights strength of negative selection across coding and non-coding annotations. *Nat. Genet.* **50**, 1600–1607 (2018).
28. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
29. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide

association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).

30. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).
31. Dong, S. *et al.* Annotating and prioritizing human non-coding variants with RegulomeDB v.2. *Nat. Genet.* **55**, 724–726 (2023).
32. Freshour, S. L. *et al.* Integration of the Drug-Gene Interaction Database (DGIdb 4.0) with open crowdsource efforts. *Nucleic Acids Res.* **49**, (2021).
33. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016).
34. Wu, Y. *et al.* Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. *Nat. Commun.* **9**, 918 (2018).
35. Qi, T. *et al.* Genetic control of RNA splicing and its distinct role in complex trait variation. *Nat. Genet.* **54**, 1355–1363 (2022).
36. Qi, T. *et al.* Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. *Nat. Commun.* **9**, 2282 (2018).
37. Nott, A. *et al.* Brain cell type-specific enhancer-promoter interactome maps and diseaserisk association. *Science* **366**, 1134–1139 (2019).
38. Palmer, D. S. *et al.* Exome sequencing in bipolar disorder identifies AKAP11 as a risk gene shared with schizophrenia. *Nat. Genet.* **54**, 541–547 (2022).
39. Wray, N. R. *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018).
40. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
41. Cook, S. *et al.* Accurate imputation of human leukocyte antigens with CookHLA. *Nat. Commun.* **12**, 1264 (2021).
42. Lam, M. *et al.* RICOPILI: Rapid Imputation for COnsortias PIpeLIne. *Bioinformatics* **36**, 930–933 (2020).

43. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).

44. Weissbrod, O. *et al.* Leveraging fine-mapping and multipopulation training data to improve cross-population polygenic risk scores. *Nat. Genet.* **54**, 450–458 (2022).

45. Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C. A. & Smoller, J. W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.* **10**, 1776 (2019).

46. O'Connell, K. S. *et al.* Genetic diversity enhances gene discovery for bipolar disorder. *medRxiv* 2023.10.07.23296687 (2023) doi:10.1101/2023.10.07.23296687.

47. Lee, S. H., Goddard, M. E., Wray, N. R. & Visscher, P. M. A better coefficient of determination for genetic profile analysis. *Genet. Epidemiol.* **36**, 214–224 (2012).

48. Momin, M. M., Lee, S., Wray, N. R. & Lee, S. H. Significance tests for R of out-of-sample prediction using polygenic scores. *Am. J. Hum. Genet.* **110**, 349–358 (2023).

49. Ikeda, M. *et al.* A genome-wide association study identifies two novel susceptibility loci and trans population polygenicity associated with bipolar disorder. *Mol. Psychiatry* **23**, 639–647 (2018).

50. Moon, S. *et al.* The Korea Biobank Array: Design and Identification of Coding Variants Associated with Blood Biochemical Traits. *Sci. Rep.* **9**, 1382 (2019).

51. Kruth, K. A., Grisolano, T. M., Ahern, C. A. & Williams, A. J. SCN2A channelopathies in the autism spectrum of neuropsychiatric disorders: a role for pluripotent stem cells? *Mol. Autism* **11**, 23 (2020).

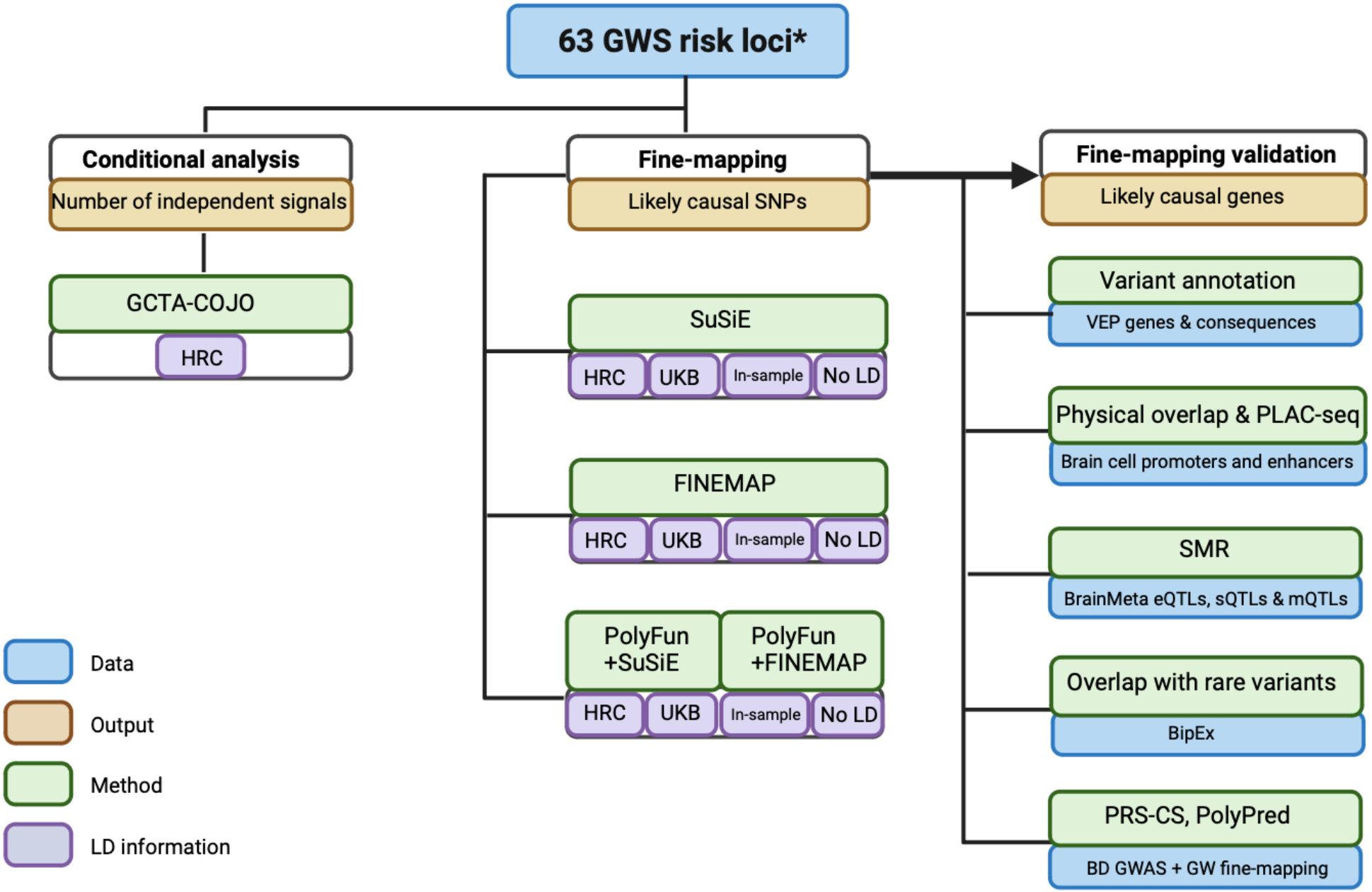
52. Kuo, M.-W., Wang, C.-H., Wu, H.-C., Chang, S.-J. & Chuang, Y.-J. Soluble THSD7A is an N-glycoprotein that promotes endothelial cell migration and tube formation in angiogenesis. *PLoS One* **6**, e29000 (2011).

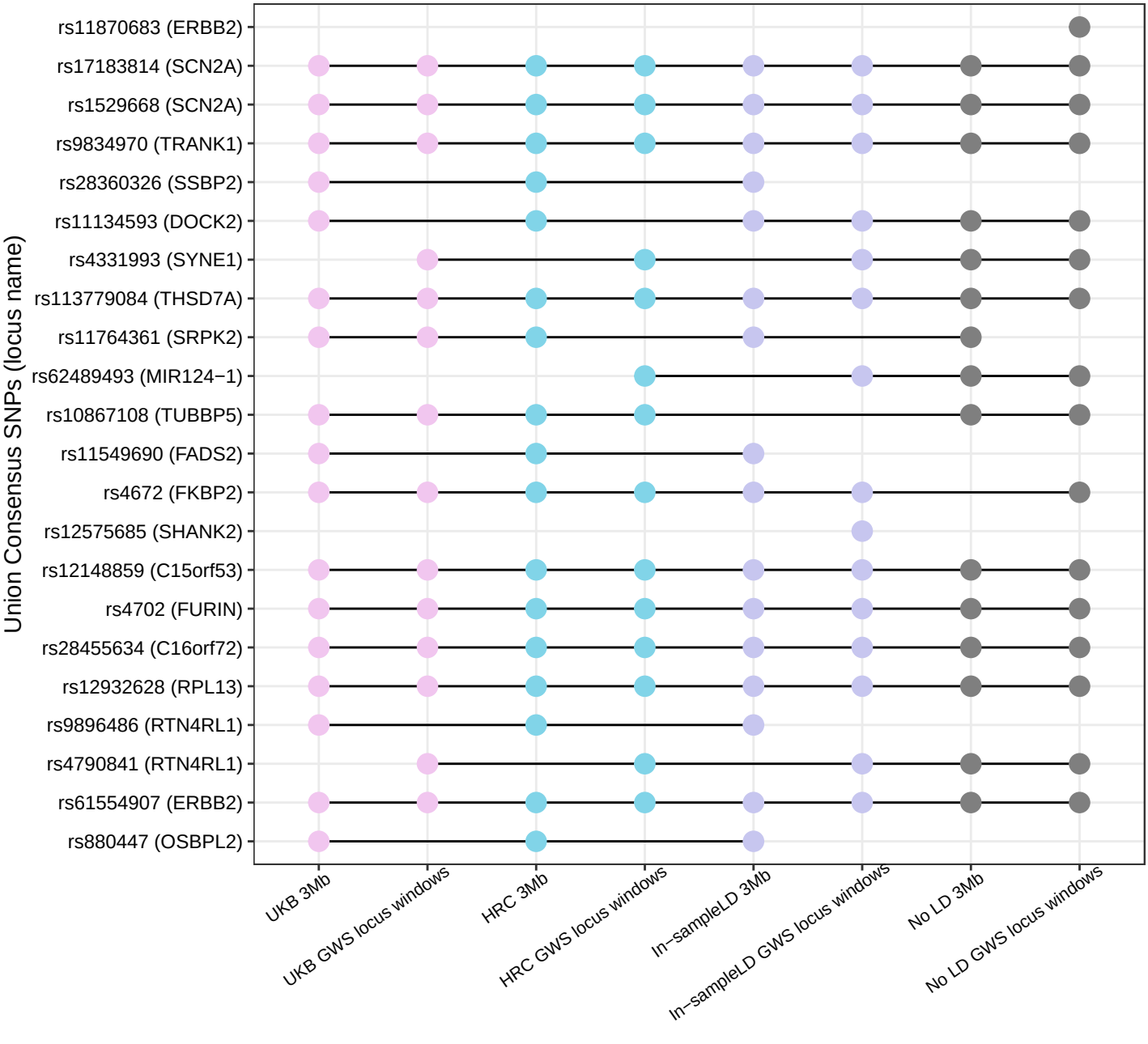
53. Ho, A. M.-C. *et al.* Mood-Stabilizing Antiepileptic Treatment Response in Bipolar Disorder: A Genome-Wide Association Study. *Clin. Pharmacol. Ther.* **108**, 1233–1242 (2020).

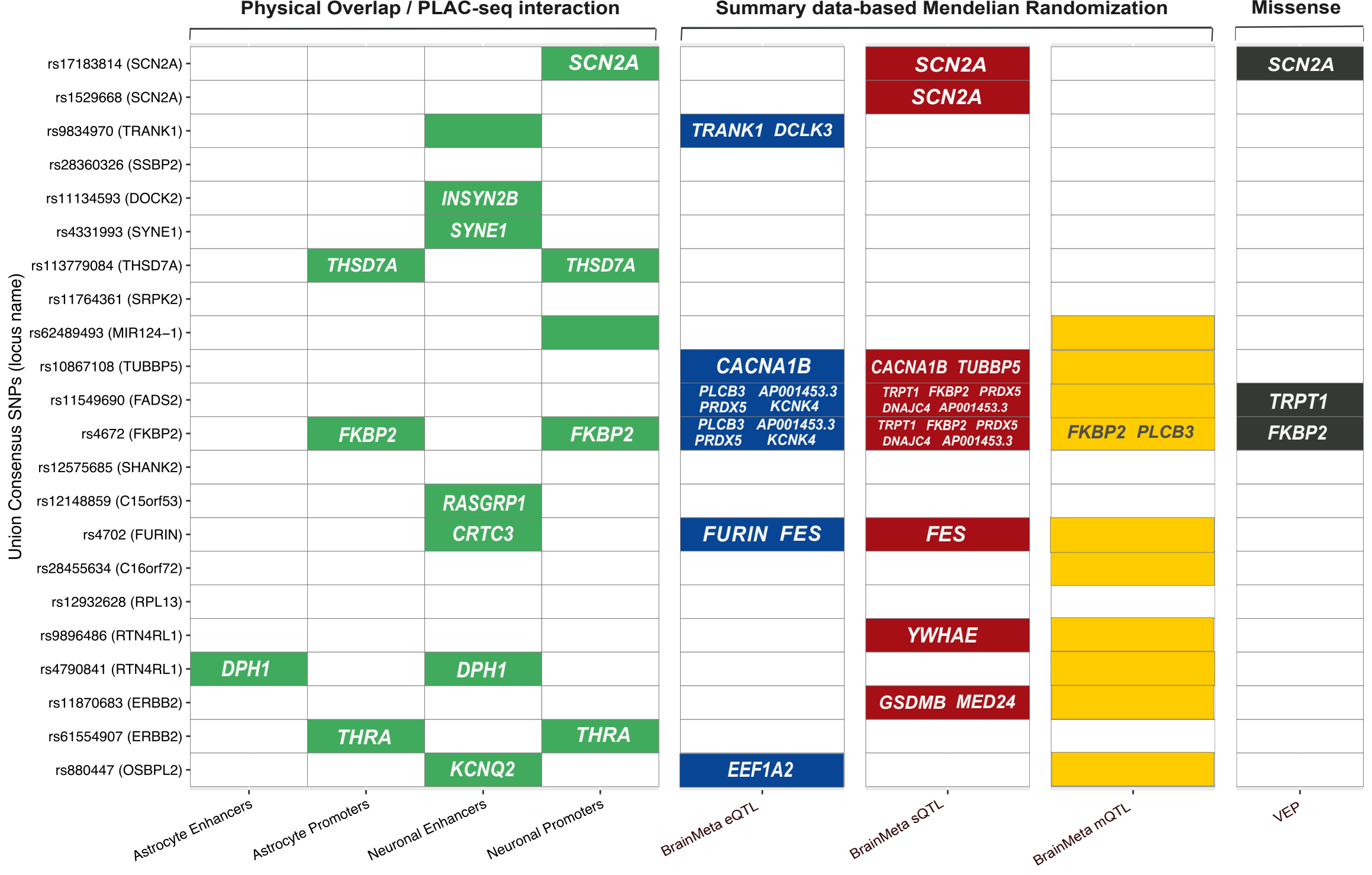
54. Schrode, N. *et al.* Synergistic effects of common schizophrenia risk variants. *Nat. Genet.*

51, 1475–1485 (2019).

55. 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).
56. Jin, Y. J. *et al.* Molecular cloning of a membrane-associated human FK506- and rapamycin-binding protein, FKBP-13. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 6677–6681 (1991).
57. Rathje, M. *et al.* Genetic variants in the bipolar disorder risk locus SYNE1 that affect CPG2 expression and protein function. *Mol. Psychiatry* **26**, 508–523 (2021).
58. Urreizti, R. *et al.* Correction: DPH1 syndrome: two novel variants and structural and functional analyses of seven missense variants identified in syndromic patients. *Eur. J. Hum. Genet.* **28**, 138 (2020).
59. Andreassen, O. A. *et al.* Genetic pleiotropy between multiple sclerosis and schizophrenia but not bipolar disorder: differential involvement of immune-related gene loci. *Mol. Psychiatry* **20**, 207–214 (2015).
60. International Schizophrenia Consortium *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).
61. Sekar, A. *et al.* Schizophrenia risk from complex variation of complement component 4. *Nature* **530**, 177–183 (2016).
62. Yuan, K. *et al.* Fine-mapping across diverse ancestries drives the discovery of putative causal variants underlying human complex traits and diseases. *medRxiv* (2023) doi:10.1101/2023.01.07.23284293.




Union Consensus SNPs (locus name)
rs11870683 (ERBB2)
rs17183814 (SCN2A)
rs1529668 (SCN2A)
rs9834970 (TRANK1)
rs28360326 (SSBP2)
rs11134593 (DOCK2)
rs4331993 (SYNE1)
rs113779084 (THSD7A)
rs11764361 (SRPK2)
rs62489493 (MIR124−1)
rs10867108 (TUBBP5)
rs11549690 (FADS2)
rs4672 (FKBP2)
rs12575685 (SHANK2)
rs12148859 (C15orf53)
rs4702 (FURIN)
rs28455634 (C16orf72)
rs12932628 (RPL13)
rs9896486 (RTN4RL1)
rs4790841 (RTN4RL1)
rs61554907 (ERBB2)
rs880447 (OSBPL2)
UKB 3Mb
UKB GWS locus windows
HRC 3Mb
HRC GWS locus windows
In−sampleLD 3Mb
In−sampleLD GWS locus windows
No LD 3Mb
No LD GWS locus windows


Physical Overlap / PLAC-seq interaction
Summary data-based Mendelian Randomization
Missense
Union Consensus SNPs (locus name)
rs17183814 (SCN2A)
rs1529668 (SCN2A)
rs9834970 (TRANK1)
rs28360326 (SSBP2)
rs11134593 (DOCK2)
rs4331993 (SYNE1)
rs113779084 (THSD7A)
rs11764361 (SRPK2)
rs62489493 (MIR124−1)
rs10867108 (TUBBP5)
rs11549690 (FADS2)
rs4672 (FKBP2)
rs12575685 (SHANK2)
rs12148859 (C15orf53)
rs4702 (FURIN)
rs28455634 (C16orf72)
rs12932628 (RPL13)
rs9896486 (RTN4RL1)
rs4790841 (RTN4RL1)
rs11870683 (ERBB2)
rs61554907 (ERBB2)
rs880447 (OSBPL2)
SCN2A
INSYN2B
SYNE1
THSD7A
RASGRP1
CRTC3
DPH1
KCNQ2
FKBP2
THRA
TRANK1 DCLK3
CACNA1B
PLCB3 AP001453.3 PRDX5 KCNK4
FURIN FES
EEF1A2
SCN2A
SCN2A
CACNA1B TUBBP5
TRPT1 FKBP2 PRDX5 DNAJC4 AP001453.3
FES
YWHAE
GSDMB MED24
FKBP2 PLCB3
SCN2A
TRPT1
FKBP2
Astrocyte Enhancers
Astrocyte Promoters
Neuronal Enhancers
Neuronal Promoters
BrainMeta eQTL
BrainMeta sQTL
BrainMeta mQTL
VEP