Title: Fine-mapping genomic loci refines bipolar disorder risk genes

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 J. Coombes[9], Jaeyoung Kim[10], Xiaoxi Liu[11,12], Chikashi Terao[11,12,13], Kevin S. O.'Connell[14,15], Mark J. Adams[16], Rolf Adolfsson[17], Martin Alda[18,19], Lars Alfredsson[20], Till F. M. Andlauer[21], Ole A. Andreassen[14,15], Anastasia Antoniou[22], Bernhard T. Baune[23,24,25], Susanne Bengesser[26], Joanna Biernacka[9,27], Michael Boehnke[28], Rosa Bosch[29,30], Murray J. Cairns[31], Vaughan J. Carr[32], Miquel Casas[29,30], Stanley Catts[33], Sven Cichon[34,35,36,37], Aiden Corvin[38], Nicholas Craddock[39], Konstantinos Dafnas[22], Nina Dalkner[26], Udo Dannlowski[40], Franziska Degenhardt[35,41], Arianna Di Florio[39,42], Dimitris Dikeos[43], Frederike Tabea Fellendorf[26], Panagiotis Ferentinos[43,44], Andreas J. Forstner[35,37,45], Liz Forty[39], Mark Frye[27], Janice M. Fullerton[46,47], Micha Gawlik[48], Ian R. Gizer[49], Katherine Gordon-Smith[50], Melissa J. Green[46,51], Maria Grigoroiu-Serbanescu[52], José Guzman-Parra[53], Tim Hahn[40], Frans Henskens[31], Jan Hillert[54], Assen V. Jablensky[55], Lisa Jones[50], Ian Jones[39], Lina Jonsson[56], John R. Kelsoe[57], Tilo Kircher[58], George Kirov[39], Sarah Kittel-Schneider[48,59,60], Manolis Kogevinas[61], Mikael Landén[56,62], Marion Leboyer[63,64], Melanie Lenger[26], Jolanta Lissowska[65], Christine Lochner[66], Carmel Loughland[31], Donald J. MacIntyre[16], Nicholas G. Martin[67,68], Eirini Maratou[69], Carol A. Mathews[70], Fermin Mayoral[53], Susan L. McElroy[71], Nathaniel W. McGregor[72], Andrew McIntosh[16], Andrew McQuillin[73], Patricia Michie[31], Philip B. Mitchell[51], Paraskevi Moutsatsou[69], Bryan Mowry[33], Bertram Müller-Myhsok[74,75], Richard M. Myers[76], Igor Nenadić[77,78], Caroline M. Nievergelt[57], Markus M. Nöthen[35], John Nurnberger[79,80,81], Michael O.'Donovan[39], Claire O.'Donovan[18], Roel A. Ophoff[82,83,84], Michael J. Owen[39], Christos Pantelis[85], Carlos Pato[86], Michele T. Pato[86], George P. Patrinos[87,88,89,90], Joanna M. Pawlak[91], Roy H. Perlis[92,93], Evgenia Porichi[43], Danielle Posthuma[94,95], Josep Antoni Ramos-Quiroga[29,96,97,98], Andreas Reif[59], Eva Z. Reininghaus[26], Marta Ribasés[29,96,98,99], Marcella Rietschel[100], Ulrich Schall[31], Peter R. Schofield[51], Thomas G. Schulze[100,101,102,103,104], Laura Scott[28], Rodney J. Scott[31], Alessandro Serretti[105,106], Cynthia Shannon Weickert[51,107], Jordan W. Smoller[108,109,110], Maria Soler Artigas[29,96,97,98], Dan J. Stein[111], Fabian Streit[100,112,113,114], Claudio Toma[46,51,115], Paul Tooney[31], Marquis P. Vawter[116,117], Eduard Vieta[118], John B. Vincent[119], Irwin D. Waldman[120], Thomas Weickert[51,107], Stephanie H. Witt[100], Kyung Sue Hong[121], Masashi Ikeda[122], Nakao Iwata[122], Beata Świątkowska[123], Hong-Hee Won[10,124], Howard J. Edenberg[125,126], Stephan Ripke[7,127], Towfique Raj[3,5,6], Jonathan R. I. Coleman[44,128], 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 Quantitative Health Sciences, Mayo Clinic, Rochester, MN, USA. [10]Department of Digital Health, Samsung Advanced Institute for

Health Sciences and Technology (SAIHST), Sungkyunkwan University, Samsung Medical Center, Seoul, Republic of Korea. [11]Laboratory for Statistical and Translational Genetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. [12]Clinical Research Center, Shizuoka General Hospital, Shizuoka, Japan. [13]The Department of Applied Genetics, The School of Pharmaceutical Sciences, University of Shizuoka, Shizuoka, Japan. [14]Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway. [15]Centre for Precision Psychiatry, University of Oslo, Oslo, Norway. [16]Division of Psychiatry, Centre for Clinical Brain Sciences, The University of Edinburgh, Edinburgh, UK. [17]Department of Clinical Sciences, Psychiatry, Umeå, University Medical Faculty, Umeå, Sweden. [18]Department of Psychiatry, Dalhousie University, Halifax, NS, Canada. [19]National Institute of Mental Health, Klecany, Czech Republic. [20]Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden. [21]Department of Neurology, Klinikum rechts der Isar, School of Medicine, Technical University of Munich, Munich, Germany. [22]National Kapodistrian University of Athens, 2nd Department of Psychiatry, Attikon General Hospital, Athens, Greece. [23]Department of Psychiatry, University of Münster, Münster, Germany. [24]Department of Psychiatry, Melbourne Medical School, The University of Melbourne, Melbourne, VIC, Australia. [25]The Florey Institute of Neuroscience and Mental Health, The University of Melbourne, Parkville, VIC, Australia. [26]Medical University of Graz, Division of Psychiatry and Psychotherapeutic Medicine, Graz, Austria. [27]Department of Psychiatry and Psychology, Mayo Clinic, Rochester, MN, USA. [28]Center for Statistical Genetics and Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA. [29]Instituto de Salud Carlos III, Biomedical Network Research Centre on Mental Health (CIBERSAM), Madrid, Spain. [30]Programa SJD MIND Escoles, Hospital Sant Joan de Déu, Institut de Recerca Sant Joan de Déu, Esplugues de Llobregat, Spain. [31]University of Newcastle, Newcastle, NSW, Australia. [32]School of Clinical Medicine, Discipline of Psychiatry and Mental Health, University of New South Wales, Sydney, NSW, Australia. [33]University of Queensland, Brisbane, QLD, Australia. [34]Department of Biomedicine, University of Basel, Basel, Switzerland. [35]Institute of Human Genetics, University of Bonn, School of Medicine and University Hospital Bonn, Bonn, Germany. [36]Institute of Medical Genetics and Pathology, University Hospital Basel, Basel, Switzerland. [37]Institute of Neuroscience and Medicine (INM-1), Research Centre Jülich, Jülich, Germany. [38]Neuropsychiatric Genetics Research Group, Dept of Psychiatry and Trinity Translational Medicine Institute, Trinity College Dublin, Dublin, Ireland. [39]Centre for Neuropsychiatric Genetics and Genomics, Division of Psychological Medicine and Clinical Neurosciences, Cardiff University, Cardiff, UK. [40]Institute for Translational Psychiatry, University of Münster, Münster, Germany. [41]Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University Hospital Essen, University of Duisburg-Essen, Duisburg, Germany. [42]Department of Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [43]National and Kapodistrian University of Athens, 2nd Department of Psychiatry, Attikon General Hospital, Athens, Greece. [44]Social, Genetic and Developmental Psychiatry Centre, King's College London, London, UK. [45]Centre for Human Genetics, University of Marburg, Marburg, Germany. [46]Neuroscience Research Australia, Sydney, NSW, Australia. [47]School of Biomedical Sciences, Faculty of Medicine and Health, University of New South Wales, Sydney, NSW, Australia. [48]Department of Psychiatry, Psychosomatics and Psychotherapy, Center of Mental Health, University Hospital Würzburg, Würzburg, Germany. [49]Department of Psychological Sciences, University of Missouri, Columbia, MO, USA. [50]Psychological Medicine, University of Worcester, Worcester, UK. [51]Discipline of Psychiatry and Mental Health, School of Clinical Medicine, Faculty of Medicine and Health, University of New South Wales, Sydney, NSW, Australia. [52]Biometric Psychiatric Genetics Research Unit, Alexandru Obregia Clinical Psychiatric Hospital, Bucharest, Romania. [53]Mental Health Department, University Regional Hospital, Biomedicine Institute (IBIMA), Málaga, Spain. [54]Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden. [55]University of Western Australia, Nedlands, WA, Australia. [56]Institute of Neuroscience and Physiology, University of Gothenburg, Gothenburg, Sweden. [57]Department of Psychiatry, University of California San Diego, La Jolla, CA, USA. [58]Department of Psychiatry and Psychotherapy, University of Marburg, Germany. [59]Department of Psychiatry, Psychosomatic Medicine and Psychotherapy, University Hospital Frankfurt, Frankfurt am Main, Germany. [60]Department of Psychiatry and Neurobehavioural Science, University College Cork, Cork, Ireland. [61]ISGlobal, Barcelona, Spain. [62]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [63]University of Paris Est Créteil, INSERM, IMRB, Translatiol Neuropsychiatry, Créteil, France. [64]Department of Psychiatry and Addiction Medicine, Assistance Publique - Hôpitaux de Paris, Paris, France. [65]Cancer Epidemiology and Prevention, M. Sklodowska-Curie National Research Institute of Oncology, Warsaw, Poland. [66]SA MRC Unit on Risk and Resilience in Mental Disorders, Dept of Psychiatry, Stellenbosch University, Stellenbosch, South Africa. [67]Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia. [68]School of Psychology, The University of Queensland, Brisbane, QLD, Australia. [69]National and Kapodistrian University of Athens, Medical School, Clinical Biochemistry Laboratory, Attikon

General Hospital, Athens, Greece. [70]Department of Psychiatry and Genetics Institute, University of Florida, Gainesville, FL, USA. [71]Research Institute, Lindner Center of HOPE, Mason, OH, USA. [72]Systems Genetics Working Group, Department of Genetics, Stellenbosch University, Stellenbosch, South Africa. [73]Division of Psychiatry, University College London, London, UK. [74]Department of Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Munich, Germany. [75]Munich Cluster for Systems Neurology (SyNergy), Munich, Germany. [76]Hudsolpha Institute for Biotechnology, Huntsville, AL, USA. [77]Department of Psychiatry and Psychotherapy, University of Marburg, Marburg, Germany. [78]Center for Mind, Brain and Behavior (CMBB), University of Marburg and Justus Liebig University Giessen, Giessen, Germany. [79]Stark Neurosciences Research Institute, Indiana University School of Medicine. [80]Departments of Psychiatry and Medical and Molecular Genetics, Indiana University School of Medicine. [81]Indiana University School of Medicine. [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 and Biobehavioral Science, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA. [85]Department of Psychiatry, University of Melbourne, VIC, Australia. [86]Institute for Genomic Health, SUNY Downstate Medical Center College of Medicine, Brooklyn, NY, USA. [87]University of Patras, School of Health Sciences, Department of Pharmacy, Laboratory of Pharmacogenomics and Individualized Therapy, Patras, Greece. [88]United Arab Emirates University, College of Medicine and Health Sciences, Department of Genetics and Genomics, Al-Ain, Abu Dhabi, United Arab Emirates. [89]United Arab Emirates University, Zayed Center for Health Sciences, Al-Ain, Abu Dhabi, United Arab Emirates. [90]Erasmus University Medical Center, Faculty of Medicina and Health Sciences, Department of Pathology, Clinical Bioinformatics Unit, Rotterdam, The Netherlands. [91]Department of Psychiatric Genetics, Department of Psychiatry, 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]Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. [95]Department of Clinical Genetics, Amsterdam Neuroscience, Vrije Universiteit Medical Center, Amsterdam, The Netherlands. [96]Department of Mental Health, Hospital Universitari Vall d´Hebron, Barcelona, Spain. [97]Department of Psychiatry and Forensic Medicine, Universitat Autònoma de Barcelona, Barcelona, Spain. [98]Psychiatric Genetics Unit, Group of Psychiatry Mental Health and Addictions, Vall d´Hebron Research Institut (VHIR), Universitat Autònoma de Barcelona, Barcelona, Spain. [99]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. [100]Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany. [101]Institute of Psychiatric Phenomics and Genomics (IPPG), University Hospital, LMU Munich, Munich, Germany. [102]Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [103]Department of Psychiatry and Psychotherapy, University Medical Center Göttingen, Göttingen, Germany. [104]Department of Psychiatry and Behavioral Sciences, SUNY Upstate Medical University, Syracuse, NY, USA. [105]Department of Medicine and Surgery, Kore University of Enna, Italy. [106]Oasi Research Institute-IRCCS, Troina, Italy. [107]Department of Neuroscience, SUNY Upstate Medical University, Syracuse, NY, USA. [108]Stanley Center for Psychiatric Research, Broad Institute, Cambridge, MA, USA. [109]Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA. [110]Psychiatric and Neurodevelopmental Genetics Unit (PNGU), Massachusetts General Hospital, Boston, MA, USA. [111]SAMRC Unit on Risk and Resilience in Mental Disorders, Dept of Psychiatry and Neuroscience Institute, University of Cape Town, Cape Town, South Africa. [112]Department of Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany. [113]Hector Institute for Artificial Intelligence in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany. [114]German Center for Mental Health (DZPG), partner site Mannheim/Heidelberg/Ulm, Germany. [115]Centro de Biología Molecular Severo Ochoa, Universidad Autónoma de Madrid and CSIC, Madrid, Spain. [116]Functional Genomics Laboratory, School of Medicine, University of California, Irvine, California. [117]Department of Psychiatry and Human Behavior, School of Medicine, University of California, Irvin, California. [118]Clinical Institute of Neuroscience, Hospital Clinic, University of Barcelona, IDIBAPS, CIBERSAM, Barcelona, Spain. [119]Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, ON, Canada. [120]Department of Psychology, Emory University, Atlanta, GA, USA. [121]Department of Psychiatry, University of British Columbia, Vancouver, British Columbia, Canada. [122]Department of Psychiatry, Fujita Health University School of Medicine, Toyoake, Japan. [123]Department of Environmental Epidemiology, Nofer Institute of Occupational Medicine, Lodz,

Poland. [124]Samsung Advanced Institute for Health Sciences and Technology (SAIHST), Sungkyunkwan University, Samsung Medical Center, Seoul, Republic of Korea. [125]Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN, USA. [126]Department of Medical and Molecular Genetics, Indiana University, Indianapolis, IN, USA. [127]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA. [128]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 17 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 genes involved in neurotransmission and neurodevelopment including *SCN2A, TRANK1, DCLK3, INSYN2B, SYNE1, THSD7A, CACNA1B, TUBBP5, FKBP2, RASGRP1, FURIN, FES, MED24* and *THRA* among others 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 of BD polygenic risk scores across diverse populations, and present a high-throughput fine-mapping pipeline.

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[18,22]. 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[23].

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[24]. Several fine-mapped candidates had particularly strong support for their pathogenic role in schizophrenia, due to convergence with rare variant associations[24]. 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[25] 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.

Results

Fine-mapping identifies likely causal BD variants

Stepwise conditional analyses using GCTA-COJO were performed in each of the 64 PGC3 BD GWS loci (**Supplemental Table S2**), conditioning associations on their top lead SNP and any subsequent conditionally independent associations, to identify loci that contained independent signals (conditional $P < 5x10^{-6}$). This analysis supported the existence of one association signal at 62 loci (**Supplemental Table S3**), and two independent association signals within the *MSRA* locus on chromosome 8 and the *RP1-84O15.2* locus on chromosome 8 (**Supplemental 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 options (see Methods; **Figure 1**). **Figure 2** shows the number of SNPs with a PIP >0.95 and PIP >0.50 in each fine-mapping analysis, alongside the Jaccard Index of concordance in results between each pair of the 16 fine-mapping analyses, calculated based on SNPs with PIP > 0.5 and part of a 95% credible set. Jaccard Indices ranged from 0.25 to 1 (mean 0.54 (SD = 0.20)), with higher values indicating more similar fine-mapping results (**Figure 2**). A breakdown of the Jaccard Indices for analyses grouped by LD option, statistical or functional fine-mapping and fine-mapping method are provided in the **Supplemental Figure S2**.

Functional fine-mapping analyses yielded significantly more fine-mapped SNPs compared with the corresponding statistical fine-mapping analyses at PIP > 0.95 and PIP > 0.5 ($P = 6.47x10^{-4}$ and P = 0.03 respectively) (**Figure 2)**. There were no significant differences in the numbers of SNPs fine-mapped between the four LD options, between the two statistical fine-mapping methods or between the two functional fine-mapping methods. Approximately a quarter of GWS loci (N= 16) 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 (17 SNPs), but fewer met the stricter threshold of PIP >0.95 (6 SNPs) (**Figure 3**). The number of 95% credible sets per locus varied based on the fine-mapping method (**Supplemental Figure S3)**.

The smallest 95% CS per locus for every fine-mapping method and LD reference panel (**Supplemental Figure S3**) was also calculated. Approximately ⅕ (N= 10-19) or ½ (N= 32-41) of the 63 fine-mapped loci had 95% CSs with a small number of SNPs ($N_{SNPs}< 10$). The percentage of fine-mapped loci harboring 95% CSs with $N_{SNPs}< 10$ was dependent on the fine-mapping method, with FINEMAP and PolyFun+FINEMAP harboring smaller 95% CSs and SuSiE and PolyFun+SuSiE larger 95% CSs.

The union consensus set (PIP >0.5) comprised 17 SNPs (from 16 GWS loci), indicating that many of the same SNPs were prioritized regardless of which LD reference panel was used (**Figure 3**). There were 15 SNPs consistently prioritized as the likely causal variant across all LD options (**Figure 3, Supplemental Figure S4**). Notably, while rs11870683 met consensus SNP criteria, it was only prioritized using single-variant (no LD) fine-mapping, and the multi-variant fine-mapping methods were unable to resolve the signal in this locus (**Figure 3**). The distribution of SNPs with PIP >0.50 for each GWS locus across different methods and LD options is provided in the **Supplemental Figure S4**.

Variant annotation of the union consensus SNPs via VEP[26] indicated that 5 of the 17 fall in intronic regions (**Supplemental Table S4**). Two of the union consensus SNPs are missense variants: rs17183814 in *SCN2A* (CADD: 20, ClinVar benign for seizures and developmental and epileptic encephalopathy) and rs4672 in *FKBP2* (CADD: 22.5, not in ClinVar). More details about the variant annotations of the union consensus SNPs through different online databases is provided in **Supplemental Table S4**.

QTL integrative analyses and overlap with epigenomic peaks

Summary data-based Mendelian randomization (SMR)[27,28] 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)[29] and mQTL data were from the Brain-mMeta study (adult cortex or fetal brain samples in 1,160 individuals)[30]. 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 13, 57 and 40 SNP-probe pairs for eQTL, sQTL and mQTL analyses respectively. In the eQTL analyses, there were 5 union consensus SNPs with significant P_{SMR} that passed the HEIDI (heterogeneity in dependent instruments) test for 9 different genes, suggesting that their effect on BD is mediated via gene expression in the brain (**Figure 4, Supplemental Table S5**). Three 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 6 union consensus SNPs with significant P_{SMR} results, and passing the HEIDI test, implicating 11 genes (**Figure 4, Supplemental Table S5**). In the mQTL analyses, there were 20 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 4, Supplemental Table S5**).

There were 11 union consensus SNPs that physically overlapped with active enhancers or promoters of gene expression in brain cell-types[31], particularly neurons (**Figure 4**). 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 4**). Amongst the implicated target genes through enhancer-promoter interactions are *INSYN2B*, *SYNE1*, *RASGRP1*, *CRTC3*, *DPH1* and *THRA*.

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 4**, **Supplemental Figure S5**). Assuming that a single variant may act through multiple risk genes, we took the union of the prioritized genes across the different lines of evidence described above. Taken together, the data support the roles of the following 23 genes in BD: *SCN2A, TRANK1, DCLK3, INSYN2B, SYNE1, THSD7A, CACNA1B, TUBBP5, PLCB3, AP001453.3, PRDX5, KCNK4, CRTC3, TRPT1, FKBP2, DNAJC4, RASGRP1, FURIN, FES, DPH1, GSDMB, MED24 and THRA* **(Supplemental Table S6)**. **Supplemental 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 the majority of GWS loci. We assessed the high-confidence genes for evidence of rare variant associations with BD, using data from the BipEx exome sequencing study[32]. Amongst the 23 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 Open Targets platform (*SCN2A*, *CACNA1B, PRDX5, THRA, MED24, SYNE1, KCNK4, FKBP2, RASGRP1, PLCB3, DCLK3, FURIN, FES*). Detailed literature information about the biological relevance of the high-confidence genes can be found in the **Supplemental Table S6**.

Dissecting the MHC locus

In the original GWAS, the most significant SNP in the extended MHC was rs13195402 (26.4 Mb, $P = 5.8 \times 10^{-15}$) which is a missense variant in *BTN2A1*. Conditional analysis on this SNP suggested a single association signal across the extended MHC, and there were no associations between structural haplotypes of the complement component 4 genes (*C4A/C4B*) (~31.9 Mb) and BD[16]. Here, 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 classical MHC was rs1541269 (30.1 Mb, $P = 6.71 \times 10^{-12}$, LD $r^2 = 0.55$ with the original index SNP rs13195402 in

European populations)[16]. While initially some variants in *HLA* genes reached GWS (**Supplemental Table S7)**, none remained after conditioning on rs1541269, suggesting the associations were driven by LD with more strongly associated variants located upstream (**Supplemental Figure 6, Supplemental Table S8**).

Leveraging fine-mapping to improve BD polygenic risk scores

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 PGC3 BD GWAS, two East Asian cohorts, four admixed African American cohorts, and three Latino cohorts[33–35]. 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 5, Supplemental Tables S9**). Using PRS-CS, the effective sample size-weighted phenotypic variance explained on the liability scale was 12.26% in EUR ancestries, 2.41% in East Asian ancestries, 0.20% in African American ancestries and 0.28% in Latino ancestries (**Figure 5**, **Supplemental Table S10)**. Examining fine-mapping-informed PRS, SuSiE+PRS-CS or Polypred-P explained more phenotypic variance than PRS-CS in all cohorts, with PolyPred-P showing the best performance (**Figure 5**). 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) (**Supplemental Table S9, Figure 5).**

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 17 likely causal SNPs for BD in 16 genomic loci. We linked these SNPs to genes and investigated their likely functional consequences, by integrating variant annotations, brain cell-type epigenomic annotations and brain QTLs. Convergence of evidence across these analyses prioritized 23 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 and LD reference panels. This comprised 17 SNPs (from 16 GWS loci), indicating that many of the same SNPs were prioritized across fine-mapping analyses (**Figure 3**). Linking these SNPs to genes and investigating their likely functional consequences using computational approaches and relevant datasets, prioritized 23 high-

confidence genes (**Figure 4**). Overall, we hypothesized that a single putative causal SNP may influence multiple genes due to various factors such as the impact of enhancer elements on multiple genes' expression, overlap of eQTLs and sQTLs with epigenomic annotations and missense variants, and overlapping genomic coordinates of genes[29,36,37].

This study uncovered novel insights into BD. Six of the genes prioritized have synaptic functions, including two with presynaptic and four with postsynaptic annotations. The functions of these genes encompass both cellular excitability (regulation of neurotransmitter levels and membrane potential) and cellular organization (arrangement of the actin cytoskeleton, endocytosis, and the postsynaptic specialization). Prioritized genes implicate a variety of neurotransmitters, both excitatory and inhibitory. These findings highlight the impact of BD risk variants on diverse aspects of synaptic signaling. While all prioritized genes are expressed in the brain and most display enrichment of expression in several brain cell types, three of the genes prioritized have enhanced expression in cells of the gut, including gastric mucous secreting cells, and proximal and distal enterocytes. These cells play roles in intestinal permeability, inflammation and the enteric nervous system, and our findings lend genetic support to the involvement of the microbiota–gut–brain axis in BD[38]. The *PLCB3*, *KCNK4*, and *DPH1* genes prioritized have previously been linked to neurodevelopmental delay[39–41] but not BD. Our study also provides novel insights into the potential molecular mechanisms underlying known BD risk genes. For example, results suggest that fine-mapped variants impact BD through alternate splicing of *SCN2A* and *CACNA1B* in the brain*,* findings which may inform functional laboratory experiments.

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[42,43]. 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[44].

Fine-mapping-informed PRS, developed by combining GWAS effect sizes and genome-wide fine-mapping effect sizes using PolyPred, explained a greater proportion of phenotypic variance, compared with PRS based on GWAS effect sizes alone. This adds support to our fine-mapping results, as leveraging information on causal effect sizes rather than relying solely on association

statistics should improve genetic risk prediction. Under the assumption that the causal variants are shared across ancestries, we anticipated that fine-mapping-informed PRS would improve the transferability of BD PRS into diverse genetic ancestries. Indeed, there was a modest increase in the phenotypic variance explained relative to standard PRS in all genetic ancestry groups. However, the performance of all PRS in non-European cohorts still lagged greatly behind that in Europeans (**Figure 5, Supplemental Table S9, S10**), emphasizing the need for larger studies in diverse genetic ancestries and further development of methods to improve PRS transferability between ancestries.

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]. FINEMAP, using a shotgun stochastic algorithm, refines promising SNP sets efficiently by focusing on a subset with higher PIPs, making it well-suited for dense genomic data. In contrast, SuSiE's Bayesian algorithm accommodates LD structure and identifies multiple causal signals within loci, offering credible sets that increase confidence in the discovered variants. 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 (53/ 57 cohorts). 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.

While there were some differences in the number of SNPs fine-mapped (threshold of PIP > 0.5 and in a 95% credible set) by the same method using different LD options (**Figure 2**), our strategy of requiring SNPs to be fine-mapped using two methods was employed to safeguard against false positives. Moreover, although conditional analysis indicated one 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 a greater number of likely causal SNPs. As an exception, we note that one consensus SNP (rs11870683) was prioritized using single-variant (no LD) fine-mapping only, and we caution that there may be an additional or different causal SNP at this locus, since multi-variant fine-mapping methods were unable to resolve the signal. 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 PGC3 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[33] and in future, the differences in LD structure between populations could be leveraged to aid fine-mapping[45] and PRS predictions[46]. 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, we applied a conservative approach focusing on SNPs with high PIPs (>0.50), that were part of credible sets, and were supported by different fine-mapping methods. Thus, we prioritized likely causal variants or genes at 16 of the 64 GWS loci. 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.

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. 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 Supplemental Note.

Below is a list of grants supporting authors of the current work:

- Baszucki Brain Research Fund via the Milken Institute Center for Strategic Philanthropy (NM)
- US National Institute of Mental Health – PGC4: R01MH124839 (NM, OAA)
- Medical Research Foundation grant MRF-001-0012-RG-COLE-C0930 (JRIC)
- NIHR Maudsley Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London (JRIC)
- CIBER – Consorcio Centro de Investigación Biomédica en Red (CB07/09/0004) (EV)
- Instituto de Salud Carlos III, Spanish Ministry of Science and Innovation with grants PI18/00805 and PI21/00787 (integrated into the Plan Nacional de I+D+I and co- financed by ISCIII-Subdirección General de Evaluación and the Fondo Europeo de Desarrollo Regional [FEDER]) (EV)
- Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement (2021 SGR 01358) (EV)
- CERCA Programme (EV)
- Departament de Salut de la Generalitat de Catalunya for the PERIS grant SLT006/17/00357 (EV)
- European Union Horizon 2020 research and innovation program –
    - EU.3.1.1. Understanding health, wellbeing and disease: Grant No 754907 (EV)
    - EU.3.1.3. Treating and managing disease: Grant No 945151 (EV)
- CIBER – Consorcio Centro de Investigación Biomédica en Red (CB15/00154), Instituto de Salud Carlos III, Spanish Ministry of Science and Innovation and grants PI18/01788 and PI22/00464: 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) (JARQ)
- Instituto de Salud Carlos III (JARQ)
- Instituto de Salud Carlos III and NEURON BIOPHARMA, S.A: Grant Nº JTC2021 (JARQ)
- Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement: (2021 SGR 00840) (JARQ)
- European Union Horizon 2020 research and innovation program: (Eat2beNICE Grant Nº 728018; Timespan Grant Nº 965381) (JARQ)
- Australian National Health and Medical Research Council Investigator Grant (1177991 and 1176716, respectively) (PBM and PRS)

- Japan Agency for Medical Research and Development (AMED) grants: 22wm0425008, 21ek0109555, 21tm0424220, 21ck0106642, 23ek0410114, and 23tm0424225 (XL, CT, MI and NI)
- Japan Society for the Promotion of Science (JSPS) KAKENHI grants 21H02854 and JP20H00462 (XL, CT, MI and NI)
- German Research Foundation (DFG) grants FOR2107 KI588/14-1, and KI588/14-2, and KI588/20-1, KI588/22-1 to TL; DA1151/5-1, DA1151/5-2, DA1151/9-1, DA1151/10-1, DA1151/11-1 to UD; SFB-TRR393 to TL and UD.

We also gratefully acknowledge the investigators who comprise the PGC and note that 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, in accordance with the Declaration of Helsinki, with all subjects providing written informed consent.

Author contributions

All fine-mapping and QTL analyses were conducted by M.Koromina. AR replicated the results and assisted with pipeline parallelization. JH, BS and TR provided useful feedback on the fine-mapping analyses. TB, CC, XL and JK computed polygenic risk scores for the non-European cohorts. GP, AB and SR assisted with the analytical fine-mapping design of the MHC locus. NM conceived and supervised the study. The remaining authors have contributed data to the current project. MK and NM and were responsible for the primary drafting and editing of the paper. All authors reviewed the manuscript critically for important intellectual content and approved the final version of the manuscript for publication.

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. The remaining authors declare no competing interests.

Figure legends

Figure 1. Schematic workflow of the fine-mapping pipeline developed for PGC3 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 implementing single-variant (no LD) fine-mapping. PolyFun functional priors were based on the published baseline-LF2.2 UKB model[22]. 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').

Figure 2. Results and comparison of 16 fine-mapping analyses conducted. The barplot displays the number of SNPs fine-mapped with PIP > 0.5 and part of a 95% credible set on the y-axis and each fine-mapping analysis on the x-axis. The black bordered bars indicate the number of SNPs fine-mapped with PIP > 0.95 and part of a 95% credible set. Each analysis is named according to [LD option]_[fine-mapping method]. The heatmap displays the Jaccard Index of concordance in results between each pair of fine-mapping analyses, calculated based on SNPs with PIP > 0.5 and part of a 95% credible set. Jaccard Indices ranged from 0.25 to 1 (mean 0.54 (SD = 0.20)), with higher values indicating more similar fine-mapping results.

Figure 3. Plot of union consensus SNPs across all 16 fine-mapping analyses, including different LD options and fine-mapping methods. The color of the points corresponds to the LD option used: UK Biobank (pink), Haplotype Reference Consortium (blue), in-sample LD (purple) and no LD (single variant fine-mapping) (grey). Circles indicate statistical fine-mapping methods and squares indicate functional fine-mapping methods. Small shapes denote SNPs with PIP between 0.50 and 0.90, while large shapes denote SNPs with PIP above 0.95. On the x-axis, analyses are named according to [LD option] [fine-mapping method]. On the y axis, the PGC3 locus name is displayed in parenthesis after each fine-mapped SNP and indicates the name assigned to identify the locus in the original PGC3 BD GWAS publication, which is not necessarily the causal gene.

Figure 4. Summary of analyses performed to link each fine-mapped SNP to the relevant gene(s). The y-axis shows the 17 union consensus SNPs with the PGC3 locus name displayed in parenthesis after each one, which indicates the name assigned to identify the locus in the original PGC3 BD GWAS publication and not necessarily the causal gene. 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.

Figure 5. **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) with error bars indicating the 95% confidence interval around each R^2 value. 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, using the F-test for nested models.

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. 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).
23. 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.
24. Trubetskoy, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**, 502–508 (2022).
25. Mölder, F. *et al.* Sustainable data analysis with Snakemake. *F1000Res.* **10**, 33 (2021).
26. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).
27. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016).
28. Wu, Y. *et al.* Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. *Nat. Commun.* **9**, 918 (2018).
29. Qi, T. *et al.* Genetic control of RNA splicing and its distinct role in complex trait variation. *Nat. Genet.* **54**, 1355–1363 (2022).
30. Qi, T. *et al.* Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. *Nat. Commun.* **9**, 2282 (2018).
31. Nott, A. *et al.* Brain cell type-specific enhancer-promoter interactome maps and diseaserisk association. *Science* **366**, 1134–1139 (2019).
32. 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).

33. 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.

34. 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).

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

36. Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* **20**, 207–220 (2019).

37. Schoenfelder, S. & Fraser, P. Long-range enhancer-promoter contacts in gene expression control. *Nat. Rev. Genet.* **20**, 437–455 (2019).

38. Ortega, M. A. *et al.* Microbiota-gut-brain axis mechanisms in the complex network of bipolar disorders: potential clinical implications and translational opportunities. *Mol. Psychiatry* **28**, 2645–2673 (2023).

39. Bayoumi, R. *et al.* Localisation of a gene for an autosomal recessive syndrome of macrocephaly, multiple epiphyseal dysplasia, and distinctive facies to chromosome 15q26. *J. Med. Genet.* **38**, 369–373 (2001).

40. Gripp, K. W. *et al.* Syndromic disorders caused by gain-of-function variants in KCNH1, KCNK4, and KCNN3-a subgroup of K channelopathies. *Eur. J. Hum. Genet.* **29**, 1384–1395 (2021).

41. Urreizti, R. *et al.* DPH1 syndrome: two novel variants and structural and functional analyses of seven missense variants identified in syndromic patients. *Eur. J. Hum. Genet.* **28**, 64–75 (2020).

42. 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).

43. International Schizophrenia Consortium *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).

44. Sekar, A. *et al.* Schizophrenia risk from complex variation of complement component 4. *Nature* **530**, 177–183 (2016).

45. 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.

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

Methods

GWAS summary statistics and BD risk loci

Summary statistics from the latest published BD GWAS by the Psychiatric Genomics Consortium ("PGC3" study) 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 (**Supplemental Table S1**). Of these cohorts, 53 were imputed using the Haplotype Reference Consortium (HRC) EUR ancestry reference panel v1.0[47]. GWAS summary statistics were cleaned using DENTIST software[48] yielding a total of 7,598,903 SNPs. The GWAS meta-analysis identified 64 independent loci associated with BD at GWS, which were selected for fine-mapping. 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]. 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'). **Supplemental 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]. Excluding the MHC, GWS locus windows ranged between 14,960 - 3,730,000 bp in size.

Conditional analysis

Figure 1 shows an overview of the fine-mapping pipeline. First, conditional analyses were conducted using a stepwise selection procedure (--cojo-slct) via GCTA[49,50] to explore potential independent association signals within each locus, according to the LD structure of the HRC EUR reference panel. Briefly, this procedure iteratively adds SNPs to a conditional model until no conditional tests are significant (conditional $P > 5 \times 10^{-6}$)[50] to estimate the number of independent association signals per 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 (33,781 cases, 53,869 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[51]. 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 7,594,494 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. We also performed single variant fine-mapping without any LD.

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 within each GWS locus fine-mapping window. FINEMAP and SuSiE can assume multiple causal variants per locus, modeling the LD structure between them. Fine-mapping was additionally performed 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[22] to compute prior causal probabilities (priors) via an L2-regularized extension of stratified LD-score regression (S-LDSC)[52], 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[22]. Functionally-informed fine-mapping was also performed using the three LD reference panels.

In total, 16 fine-mapping analyses were conducted (12 multi-variant analyses using four fine-mapping methods and three LD reference panels and four LD-independent single-variant fine-mapping analyses), varying parameters to examine their impact and the convergence of results. We used the Jaccard Index (or Jaccard Similarity Coefficient), to summarize the concordance in the results between pairs of fine-mapping analyses. The Jaccard Index was calculated as the number of fine-mapped SNPs (PIP > 0.5 and in a 95% CS) in both fine-mapping methods (intersection), divided by the total number of fine-mapped SNPs across either method (union), and ranges from 0 (no concordance between the methods) to 1 (complete concordance between the methods). "Consensus SNPs" were defined as those in the 95% CS from at least two methods (either statistical and/or functional fine-mapping) that used the same LD option and with a PIP >0.95 or >0.50 (**Table 1**) (24 opportunities for a SNP to be a consensus SNP). The "union consensus" set of SNPs was defined as all consensus SNPs across LD options PIP >0.50, excluding SNPs identified only with the UKB LD reference panel. The numbers of SNPs fine-mapped at PIP >0.95 and PIP >0.50 between different methods and different LD options were compared using two-sided paired t-tests.

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[25], 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).

Impact of LD options and locus windows on fine-mapping

We aimed to investigate the impact of using an LD reference panel for fine-mapping or performing single variant fine-mapping with no LD, compared with using LD information calculated from the original GWAS data. The latter is typically considered the gold-standard approach, however is difficult in practice due to data availability and sharing restrictions. We performed several comparative analyses, including calculating Jaccard Indices and correlation of PIP values for fine-mapped SNPs, found that the HRC reference panel, a panel that closely resembles the genetic ancestry of the GWAS, achieves comparable fine-mapping resolution with in-sample LD estimates (**Supplemental Note)**. We also compared results from fine-mapping the GWS locus windows versus fixed 3Mb windows, which indicated substantial differences between them, and that the GWS locus windows best represent the GWS association signals from the original GWAS (**Supplemental Note)**.

Annotation of union consensus SNPs

Union consensus SNPs were characterized using the Variant Effect Predictor (VEP) (GRCh37) Ensembl release 109[26]. 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[26]. 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[53]. 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[53]. 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 the Open Targets platform[54] to detect 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)[27,28]. 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[29]. Data on methylation quantitative trait loci (mQTLs) were obtained from the Brain-mMeta study[30], 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[29]. Of the 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[29], 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[27,28]. A Bonferroni correction was applied for 13 tests in the eQTL ($P_{SMR} < 3.84 \times 10^{-3}$), 57 tests in the sQTL ($P_{SMR} < 8.77 \times 10^{-4}$) and 40 tests in the mQTL analyses ($P_{SMR} < 1.25 \times 10^{-3}$). The significance threshold for the HEIDI test (heterogeneity in dependent instruments) was $P_{HEIDI} \geq 0.01$[28]. The HEIDI test is used to identify potential violations of the Mendelian Randomization assumptions, 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[31]. 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[31] 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[32]. Using the BipEx browser[32], 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[55]. 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[56] with HLA alleles determined via sequencing. This reference was obtained from the CookHLA GitHub repository[57] (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)[58]. RICOPILI was used to perform association analysis, under an additive logistic regression model in PLINK v1.90[51], 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[59].

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[46], a method which combines effect sizes from fine-mapping with those from a standard PRS approach, such as PRS-CS[60]. PRS were calculated for individuals in 12 testing cohorts of BD cases and controls that were independent of the PGC3 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] (**Supplemental Note**).

An analytical workflow outlining the steps of the PolyPred pipeline that we followed is shown in **Supplemental 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[60]. 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[51] 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 as necessary to control for genetic ancestry[33]. 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[61], using the r2redux R package[62], 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)[62]. In addition, we computed the effective sample size-weighted combined R^2 values from PRS across different ancestries. Specifically, we transformed each R^2 to a correlation coefficient, applied the Fisher Z transformation, computed the effective sample size (Neff)-weighted mean of the Fisher Z values, and then back-transformed to obtain a combined R^2.

Data availability

GWAS data were retrieved from (Mullins et al., 2021) from the following Figshare link: https://figshare.com/articles/dataset/PGC3_bipolar_disorder_GWAS_summary_statistics/14102594. The PGC's policy is to make genome-wide summary results public. All results are made available through the Figshare open access repository at the following DOI links: (https://doi.org/10.6084/m9.figshare.27871677.v2, https://doi.org/10.6084/m9.figshare.27880524.v1, https://doi.org/10.6084/m9.figshare.27886110.v1). Data provided include MHC fine-mapping analyses of the PGC3 BIP study, as well as aggregated fine-mapping results using various methods (PolyFun+SuSiE, PolyFun+FINEMAP, SuSiE, FINEMAP) across four LD reference panels (UKB, HRC, LD, noLD) and GWS locus windows, provided in both .txt.gz and .merged.csv formats. Additional files include genome-wide fine-mapping results from SuSiE and PRS-CS protocols, and a detailed Excel file on credible sets for 12 fine-mapping analyses, specifying the SNPs and loci involved (https://doi.org/10.6084/m9.figshare.28027706.v1).

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.
Additional annotations were retrieved from the following databases: gnomAD database v4.0.0 (https://gnomad.broadinstitute.org), CADD (https://cadd.gs.washington.edu/) and ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/).

Code availability

Analysis scripts are available online at [Github: https://github.com/mkoromina/SAFFARI]. Additional scripts to recreate the visuals/graphs are available online at [Github: https://github.com/Mullins-Lab/Post-finemap_processing/]. Other software used include DENTIST [Github: https://github.com/Yves-CHEN/DENTIST], PolyPred [Github: https://github.com/omerwe/polyfun/wiki/6.-Trans-ethnic-polygenic-risk-prediction-with-PolyPred], PRS-CS [Github: https://github.com/getian107/PRScs], r2redux [Github: https://github.com/mommy003/r2redux] and RICOPILI [Github: https://github.com/Ripkelab/ricopili]. All software used is publicly available at the URLs or references cited.

Methods-only references

47. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).

48. Chen, W. *et al.* Improved analyses of GWAS summary statistics by reducing data heterogeneity and errors. *Nat. Commun.* **12**, 7117 (2021).

49. 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).

50. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* **44**, 369–75, S1–3 (2012).

51. 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).

52. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).

53. Dong, S. *et al.* Annotating and prioritizing human non-coding variants with RegulomeDB v.2. *Nat. Genet.* **55**, 724–726 (2023).

54. Ochoa, D. *et al.* The next-generation Open Targets Platform: reimagined, redesigned, rebuilt. *Nucleic Acids Res.* **51**, D1353–D1359 (2023).

55. 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).

56. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

57. Cook, S. *et al.* Accurate imputation of human leukocyte antigens with CookHLA. *Nat. Commun.* **12**, 1264 (2021).

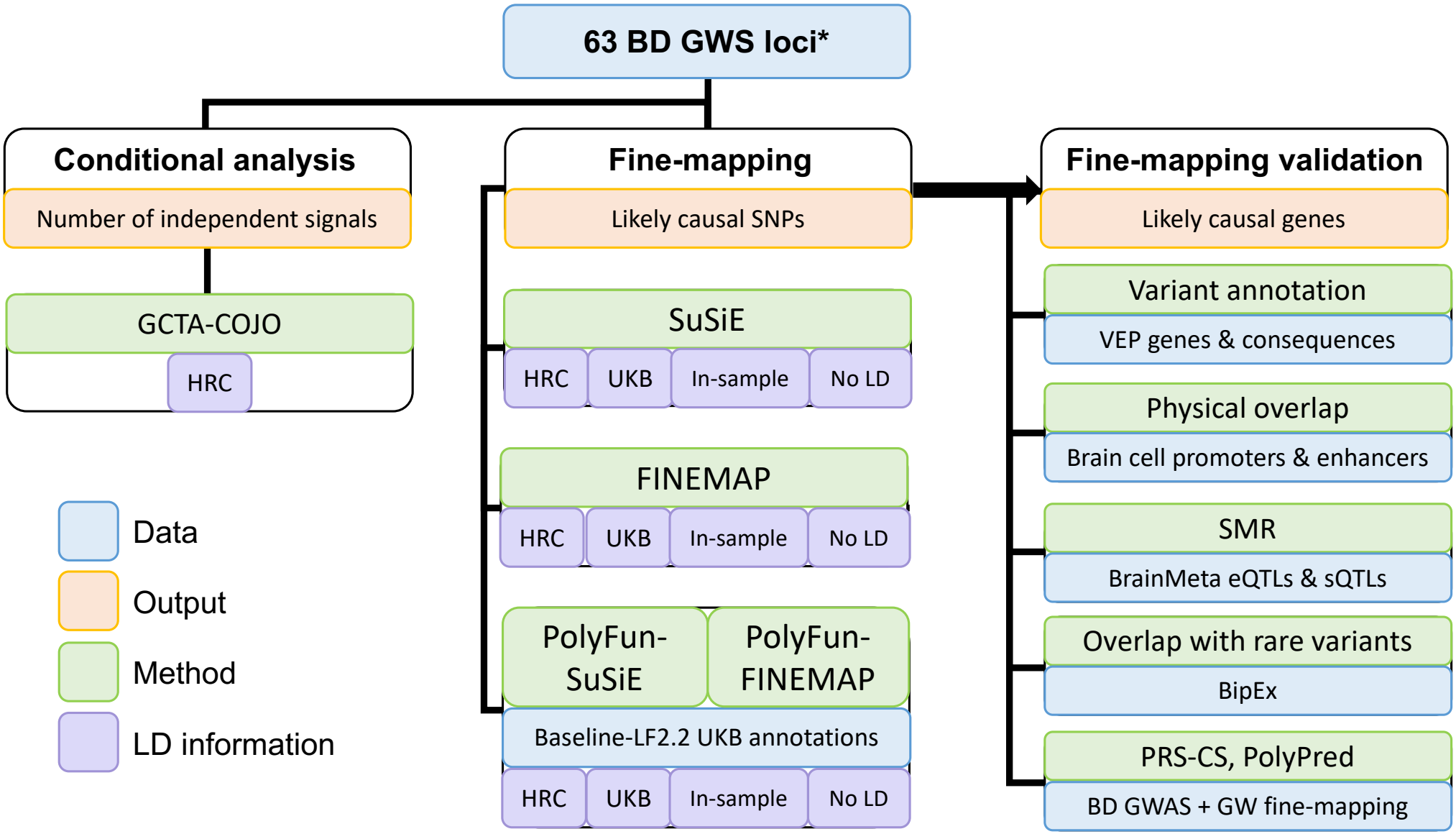
58. Lam, M. *et al.* RICOPILI: Rapid Imputation for COnsortias PIpeLIne. *Bioinformatics* **36**, 930–933 (2020).

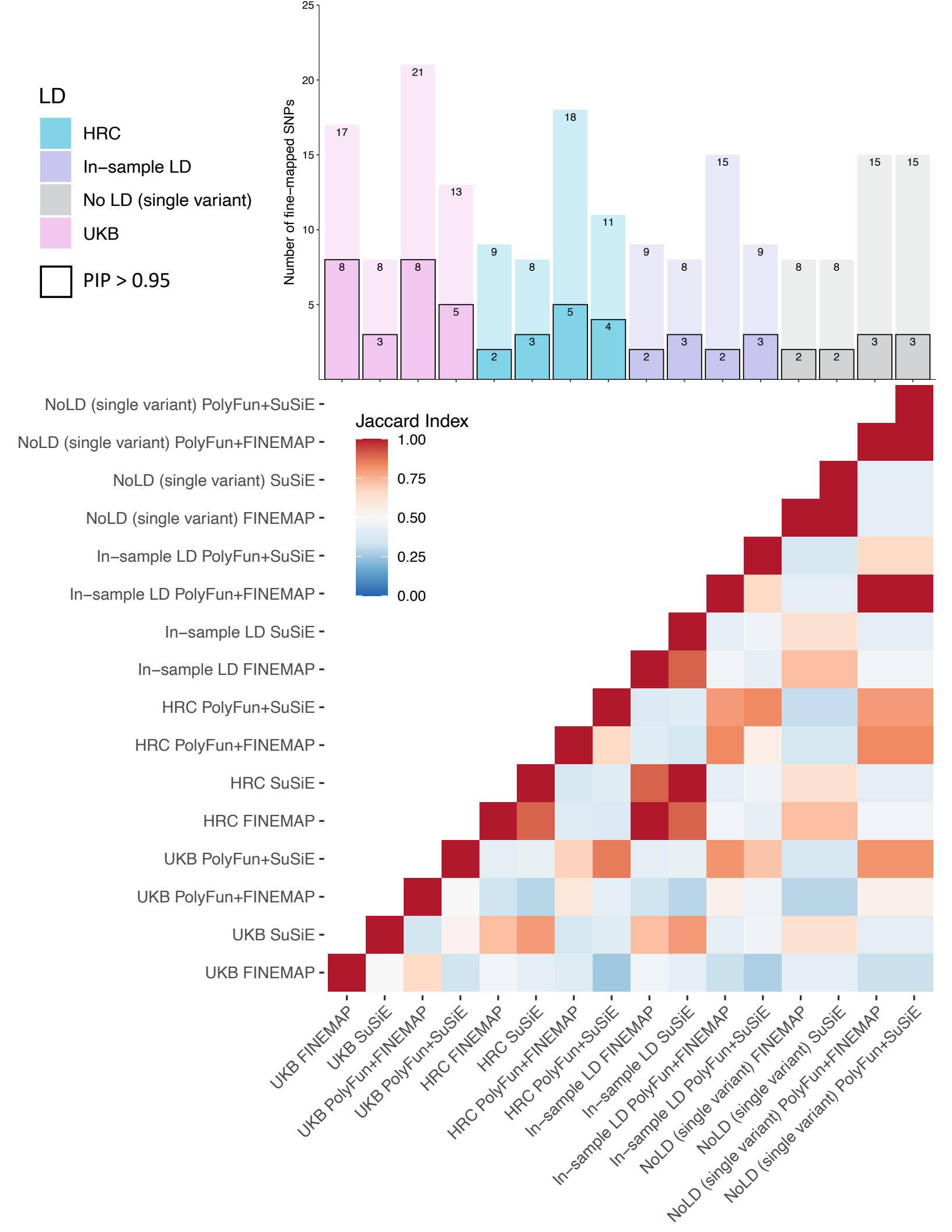
59. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).

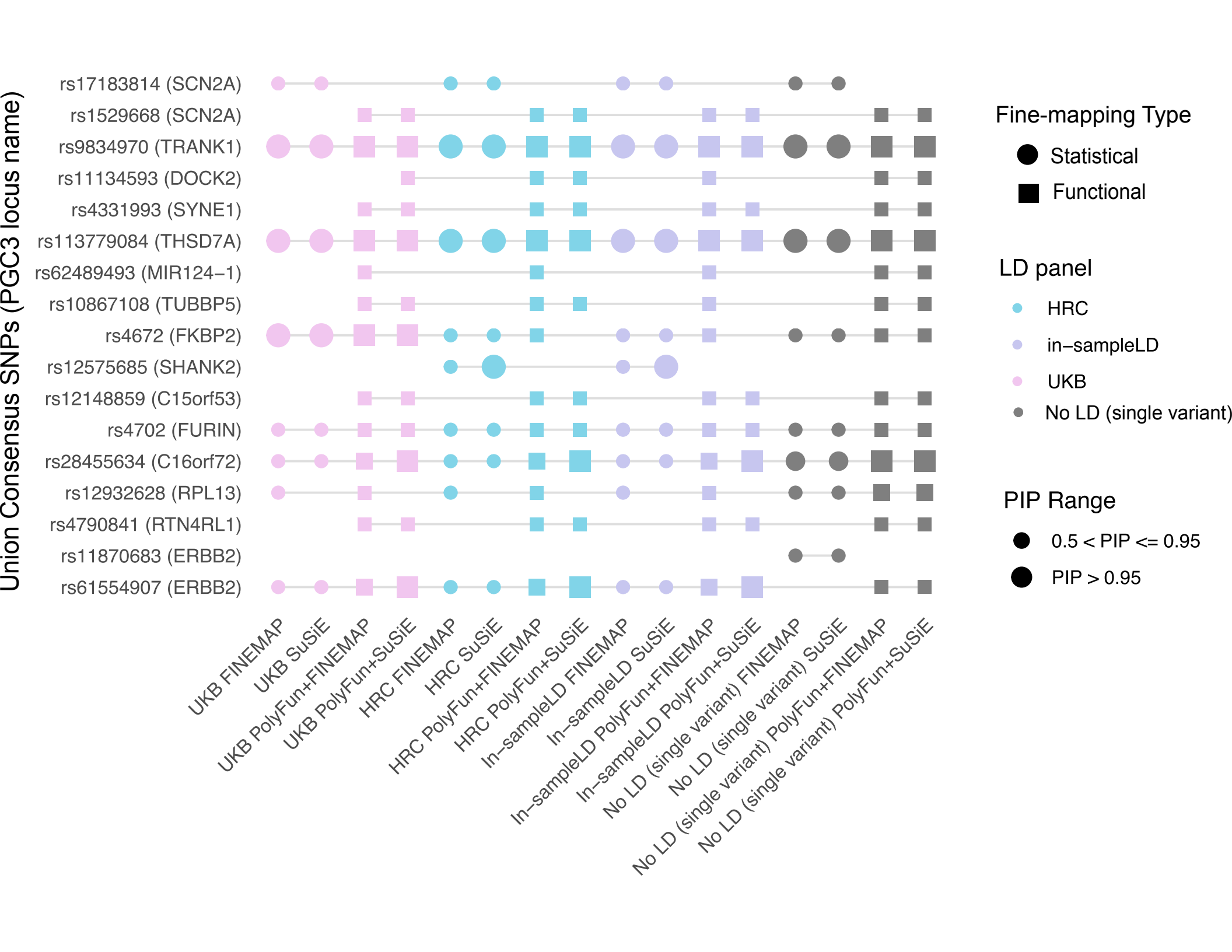
60. Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C. A. & Smoller, J. W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.* **10**, 1776 (2019).

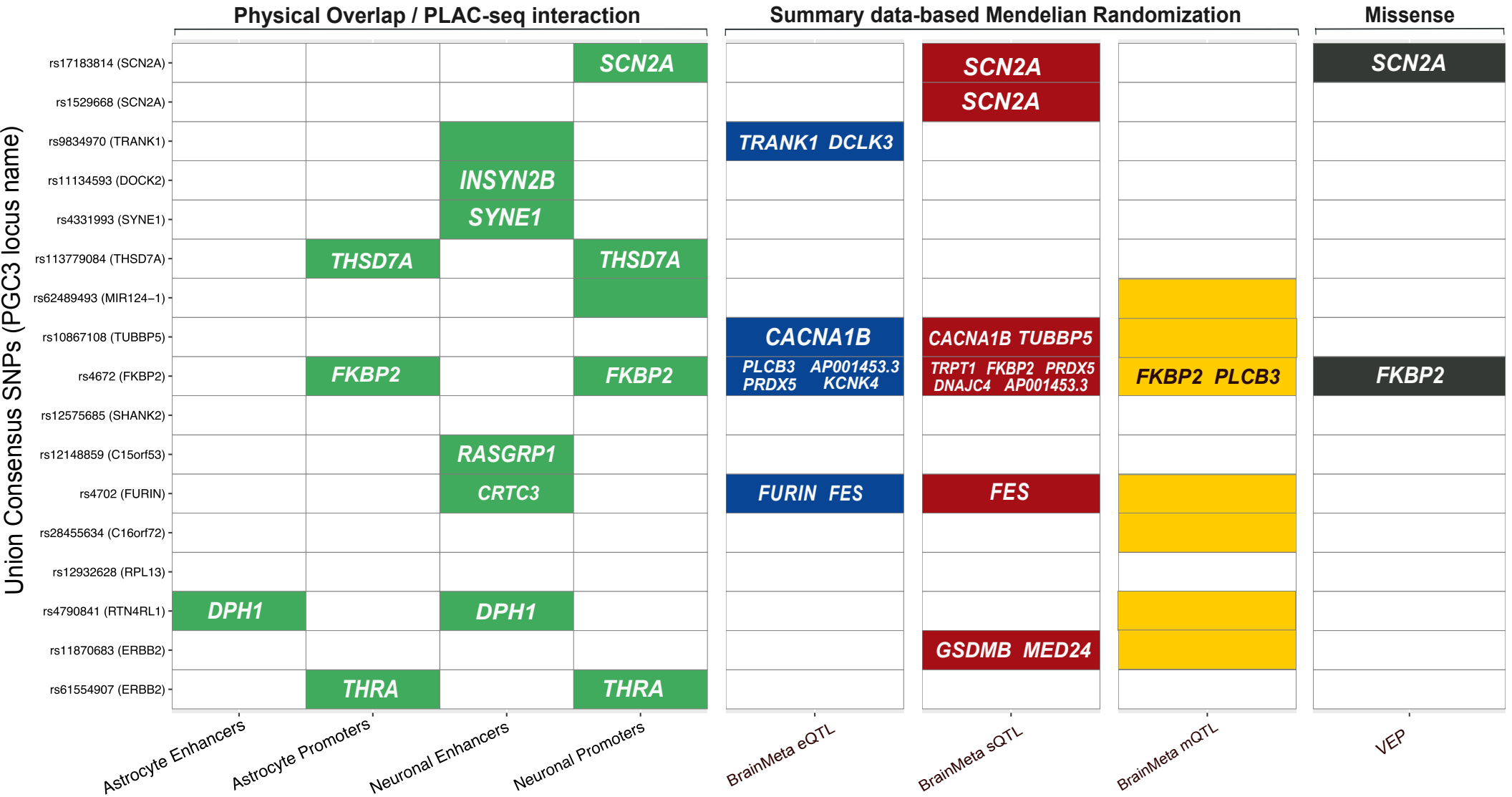
61. Lee, S. H., Goddard, M. E., Wray, N. R. & Visscher, P. M. A better coefficient of determination for genetic profile analysis. *Genet. Epidemiol.* **36**, 214–224 (2012).

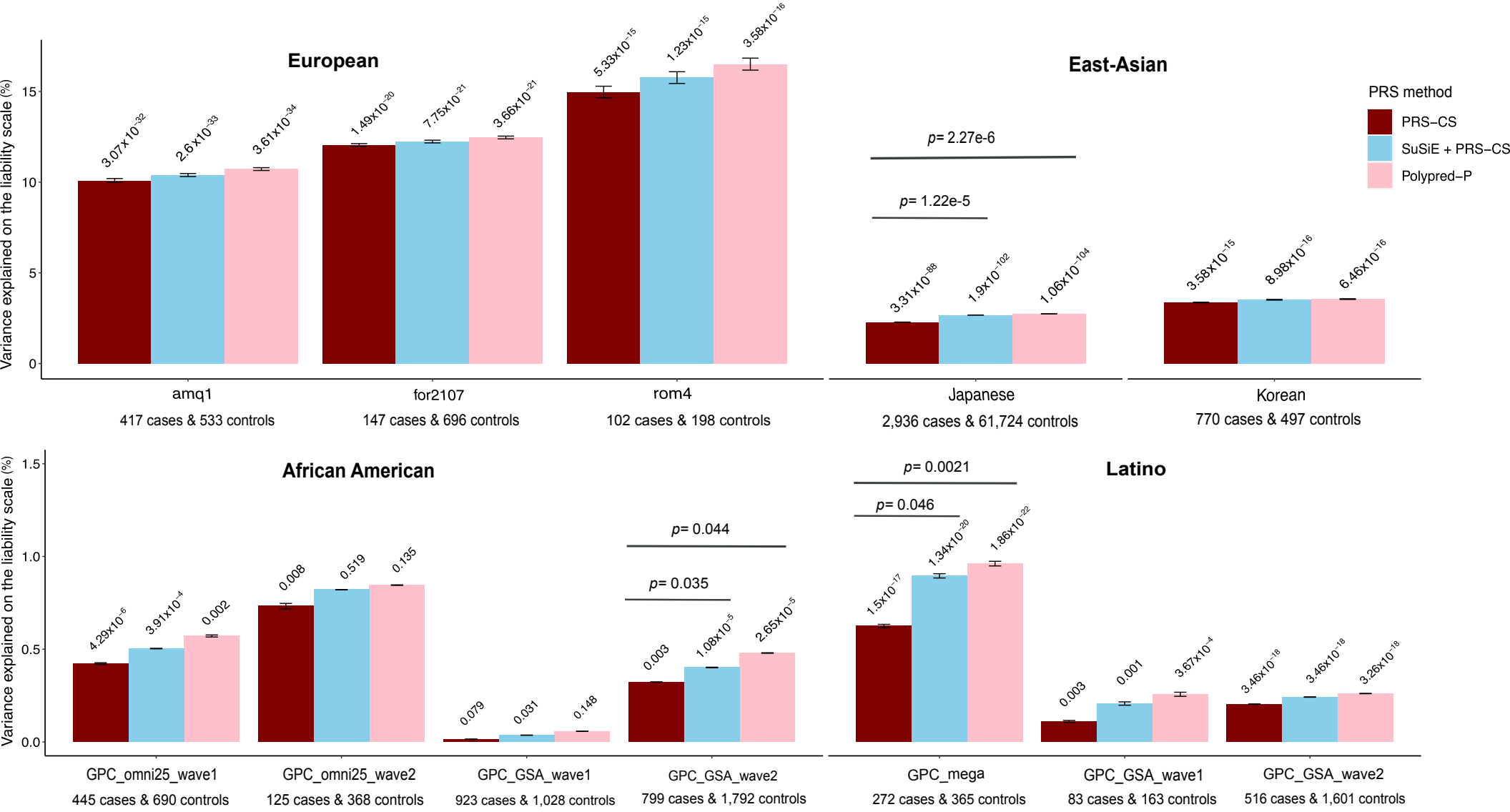
62. Momin, M. M., Lee, S., Wray, N. R. & Lee, S. H. Significance tests for R of out-of-sample prediction using polygenic scores. *Am. J. Hum. Genet.* **110**, 349–358 (2023).


63 BD GWS loci*
Conditional analysis
Number of independent signals
GCTA-COJO
HRC
Fine-mapping
Likely causal SNPs
SuSiE
HRC
UKB
In-sample
No LD
FINEMAP
HRC
UKB
In-sample
No LD
PolyFun-SuSiE
PolyFun-FINEMAP
Baseline-LF2.2 UKB annotations
HRC
UKB
In-sample
No LD
Fine-mapping validation
Likely causal genes
Variant annotation
VEP genes & consequences
Physical overlap
Brain cell promoters & enhancers
SMR
BrainMeta eQTLs & sQTLs
Overlap with rare variants
BipEx
PRS-CS, PolyPred
BD GWAS + GW fine-mapping
Data
Output
Method
LD information


LD
HRC
In−sample LD
No LD (single variant)
UKB
PIP > 0.95
Number of fine−mapped SNPs
25
20
15
10
5
17
8
8
3
21
8
8
5
13
5
9
2
8
3
18
5
11
4
9
2
8
3
15
2
9
3
8
2
8
2
15
3
15
3
Jaccard Index
1.00
0.75
0.50
0.25
0.00
NoLD (single variant) PolyFun+SuSiE
NoLD (single variant) PolyFun+FINEMAP
NoLD (single variant) SuSiE
NoLD (single variant) FINEMAP
In−sample LD PolyFun+SuSiE
In−sample LD PolyFun+FINEMAP
In−sample LD SuSiE
In−sample LD FINEMAP
HRC PolyFun+SuSiE
HRC PolyFun+FINEMAP
HRC SuSiE
HRC FINEMAP
UKB PolyFun+SuSiE
UKB PolyFun+FINEMAP
UKB SuSiE
UKB FINEMAP
UKB FINEMAP
UKB SuSiE
UKB PolyFun+FINEMAP
UKB PolyFun+SuSiE
HRC FINEMAP
HRC SuSiE
HRC PolyFun+FINEMAP
HRC PolyFun+SuSiE
In−sample LD FINEMAP
In−sample LD SuSiE
In−sample LD PolyFun+FINEMAP
In−sample LD PolyFun+SuSiE
NoLD (single variant) FINEMAP
NoLD (single variant) SuSiE
NoLD (single variant) PolyFun+FINEMAP
NoLD (single variant) PolyFun+SuSiE


Union Consensus SNPs (PGC3 locus name)
rs17183814 (SCN2A)
rs1529668 (SCN2A)
rs9834970 (TRANK1)
rs11134593 (DOCK2)
rs4331993 (SYNE1)
rs113779084 (THSD7A)
rs62489493 (MIR124−1)
rs10867108 (TUBBP5)
rs4672 (FKBP2)
rs12575685 (SHANK2)
rs12148859 (C15orf53)
rs4702 (FURIN)
rs28455634 (C16orf72)
rs12932628 (RPL13)
rs4790841 (RTN4RL1)
rs11870683 (ERBB2)
rs61554907 (ERBB2)
UKB FINEMAP
UKB SuSiE
UKB PolyFun+FINEMAP
UKB PolyFun+SuSiE
HRC FINEMAP
HRC SuSiE
HRC PolyFun+FINEMAP
HRC PolyFun+SuSiE
In−sampleLD FINEMAP
In−sampleLD SuSiE
In−sampleLD PolyFun+FINEMAP
In−sampleLD PolyFun+SuSiE
No LD (single variant) FINEMAP
No LD (single variant) SuSiE
No LD (single variant) PolyFun+FINEMAP
No LD (single variant) PolyFun+SuSiE
Fine-mapping Type
Statistical
Functional
LD panel
HRC
in−sampleLD
UKB
No LD (single variant)
PIP Range
0.5 < PIP <= 0.95
PIP > 0.95


Physical Overlap / PLAC-seq interaction
Summary data-based Mendelian Randomization
Missense
Union Consensus SNPs (PGC3 locus name)
rs17183814 (SCN2A)
rs1529668 (SCN2A)
rs9834970 (TRANK1)
rs11134593 (DOCK2)
rs4331993 (SYNE1)
rs113779084 (THSD7A)
rs62489493 (MIR124−1)
rs10867108 (TUBBP5)
rs4672 (FKBP2)
rs12575685 (SHANK2)
rs12148859 (C15orf53)
rs4702 (FURIN)
rs28455634 (C16orf72)
rs12932628 (RPL13)
rs4790841 (RTN4RL1)
rs11870683 (ERBB2)
rs61554907 (ERBB2)
SCN2A
INSYN2B
SYNE1
THSD7A
THSD7A
FKBP2
FKBP2
RASGRP1
CRTC3
DPH1
DPH1
THRA
THRA
TRANK1 DCLK3
CACNA1B
PLCB3 AP001453.3
PRDX5 KCNK4
FURIN FES
SCN2A
SCN2A
CACNA1B TUBBP5
TRPT1 FKBP2 PRDX5
DNAJC4 AP001453.3
FES
GSDMB MED24
FKBP2 PLCB3
SCN2A
FKBP2
Astrocyte Enhancers
Astrocyte Promoters
Neuronal Enhancers
Neuronal Promoters
BrainMeta eQTL
BrainMeta sQTL
BrainMeta mQTL
VEP


European
East-Asian
African American
Latino
PRS method
PRS−CS
SuSiE + PRS−CS
Polypred−P
Variance explained on the liability scale (%)
p= 2.27e-6
p= 1.22e-5
p= 0.044
p= 0.035
p= 0.0021
p= 0.046
3.07x10−32
2.6x10−33
3.61x10−34
1.49x10−20
7.75x10−21
3.66x10−21
5.33x10−15
1.23x10−15
3.58x10−16
3.31x10−88
1.9x10−102
1.06x10−104
3.58x10−15
8.98x10−16
6.46x10−16
4.29x10−6
3.91x10−4
0.002
0.008
0.519
0.135
0.079
0.031
0.148
0.003
1.08x10−5
2.65x10−5
1.5x10−17
1.34x10−20
1.86x10−22
0.003
0.001
3.67x10−4
3.46x10−18
3.46x10−18
3.26x10−18
amq1
417 cases & 533 controls
for2107
147 cases & 696 controls
rom4
102 cases & 198 controls
Japanese
2,936 cases & 61,724 controls
Korean
770 cases & 497 controls
GPC_omni25_wave1
445 cases & 690 controls
GPC_omni25_wave2
125 cases & 368 controls
GPC_GSA_wave1
923 cases & 1,028 controls
GPC_GSA_wave2
799 cases & 1,792 controls
GPC_mega
272 cases & 365 controls
GPC_GSA_wave1
83 cases & 163 controls
GPC_GSA_wave2
516 cases & 1,601 controls

Supplementary Note - Fine-mapping genomic loci refines bipolar disorder risk genes

Contents

Supplementary Figures

Figure S1. Workflow for computing standard and fine-mapping-informed polygenic risk scores (PRS). The base GWAS is the published bipolar disorder (BD) GWAS by the Psychiatric Genomics Consortium (PGC). There were 12 target 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.

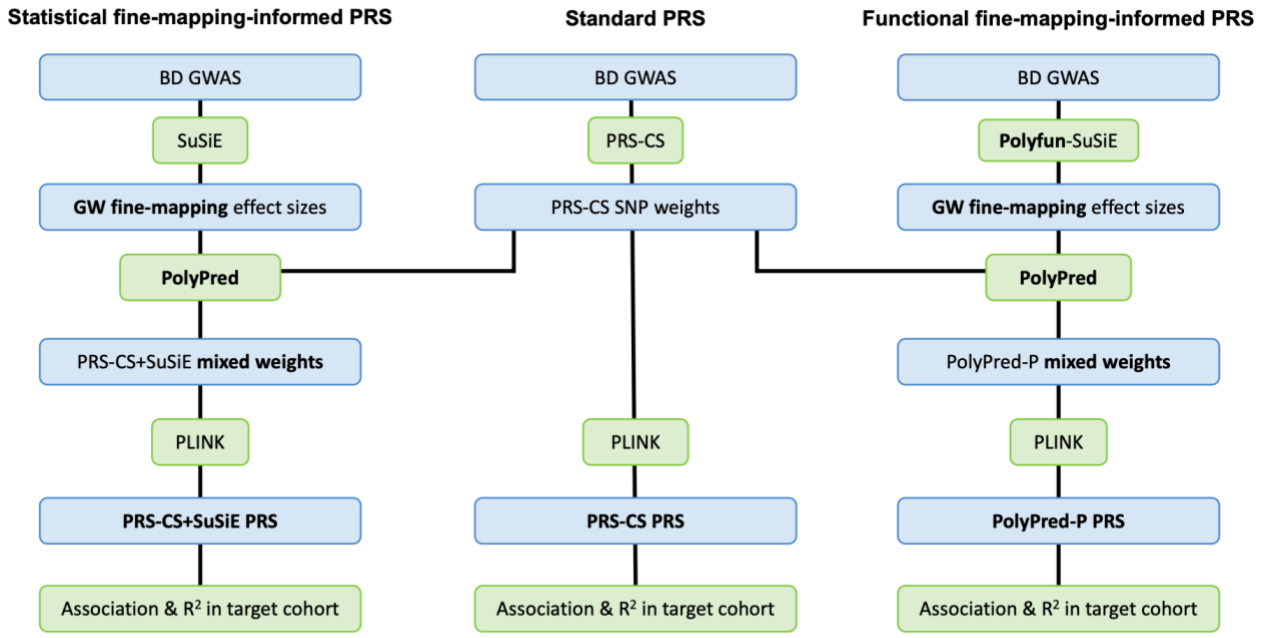
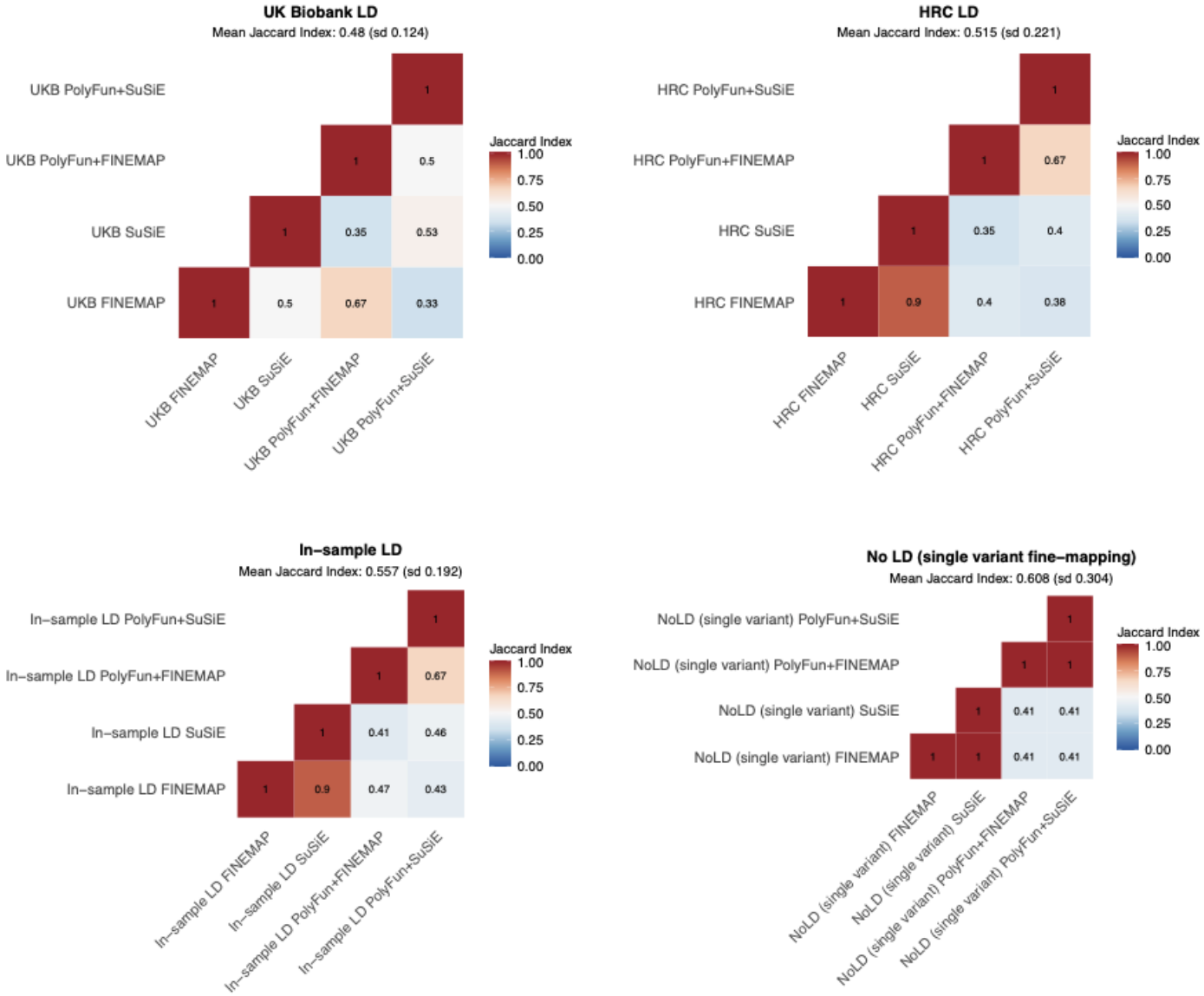
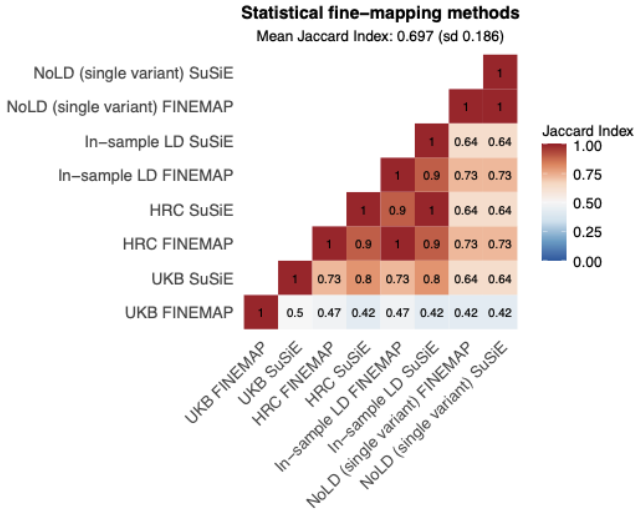


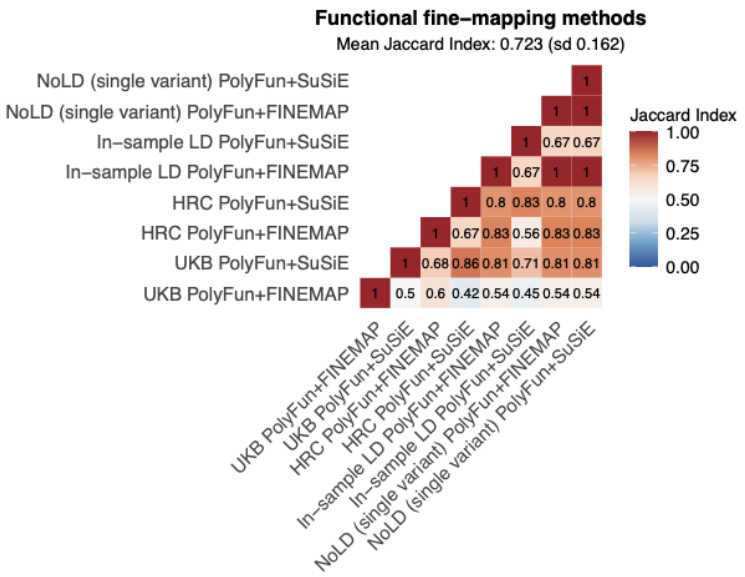
Figure S2 A-C Heatmaps of Jaccard indices for analyses grouped by A) LD option, B) statistical or functional fine-mapping and C) fine-mapping method. Jaccard Indices are calculated based on fine-mapped SNPs with PIP > 0.50 and part of 95% credible sets. The mean Jaccard Index and standard deviation (excluding the diagonal elements) is shown above each heatmap. Analyses are named according to [LD option] [fine-mapping method].

A) Jaccard Indices for analyses grouped by LD option

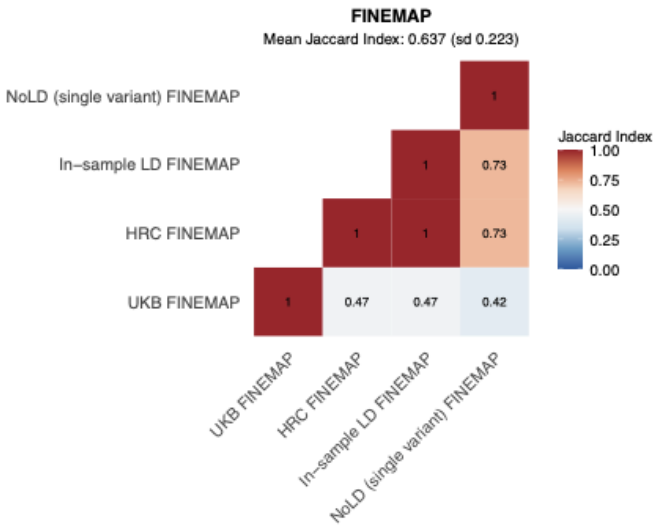


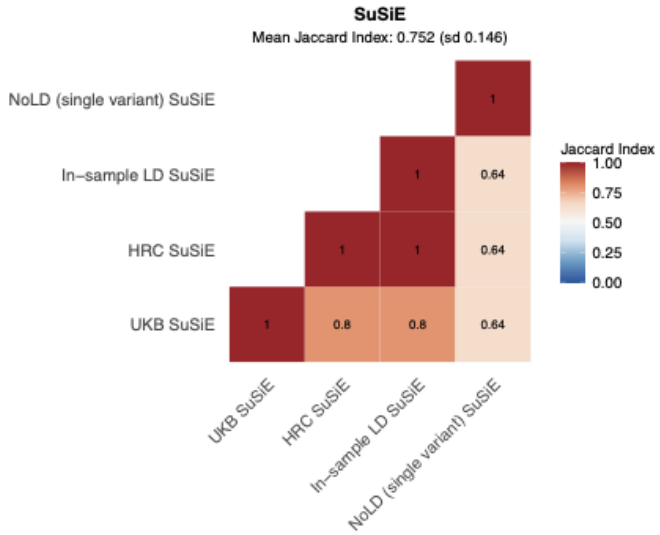
B) Jaccard Indices for analyses grouped by statistical and functional methods

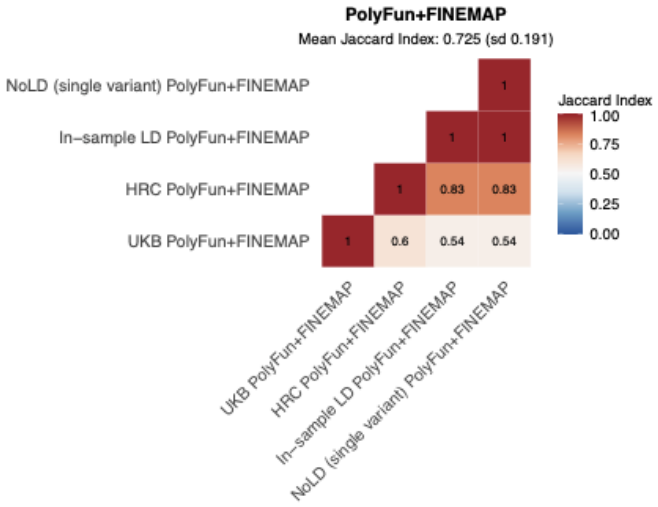




C) Jaccard Indices for analyses grouped by fine-mapping method







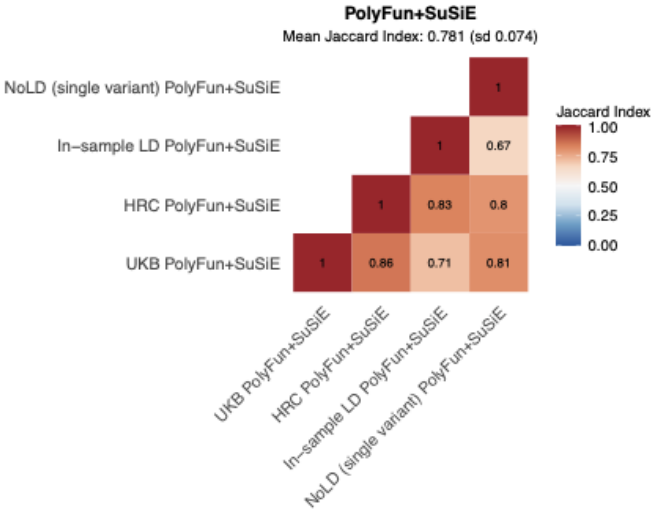
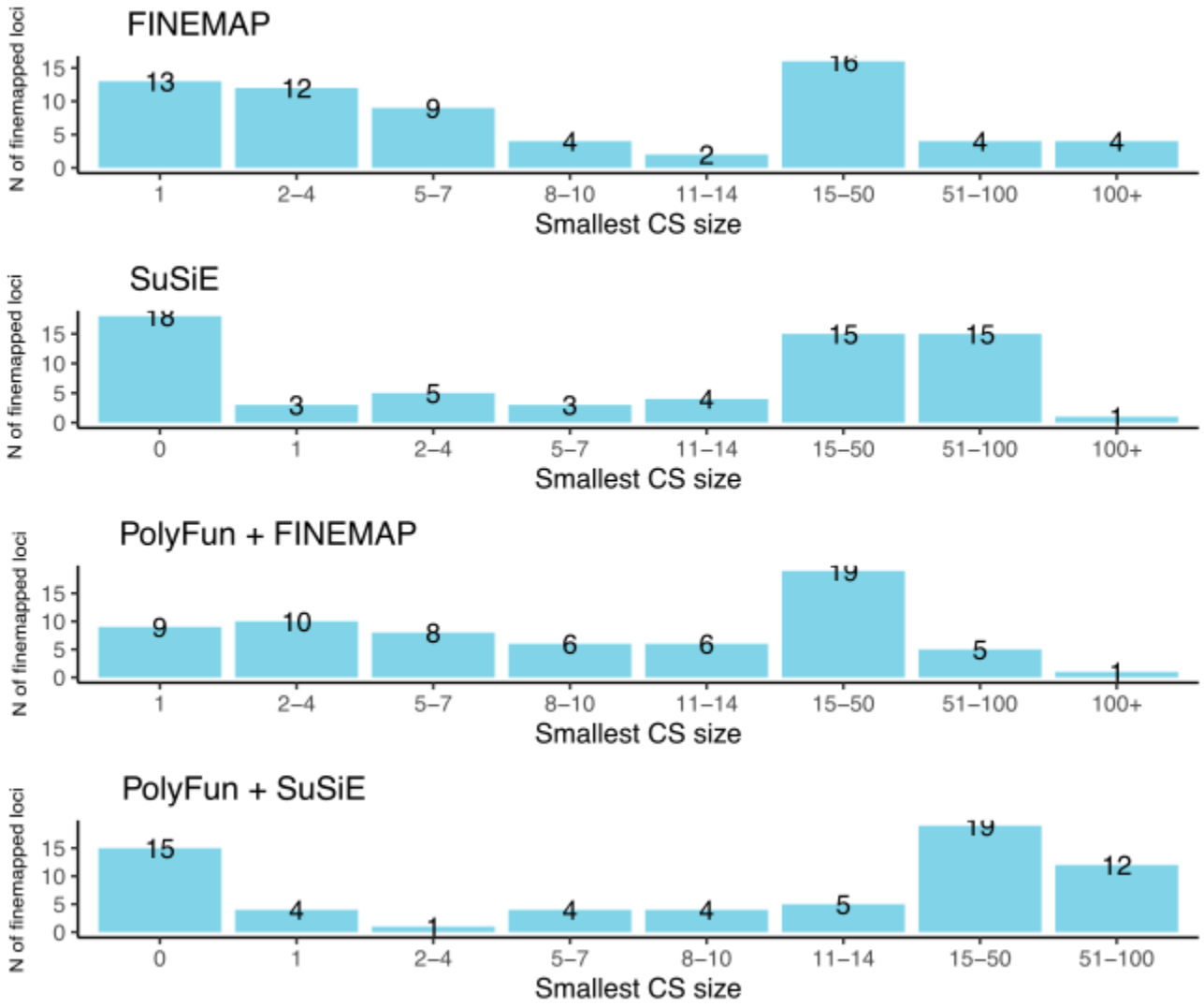
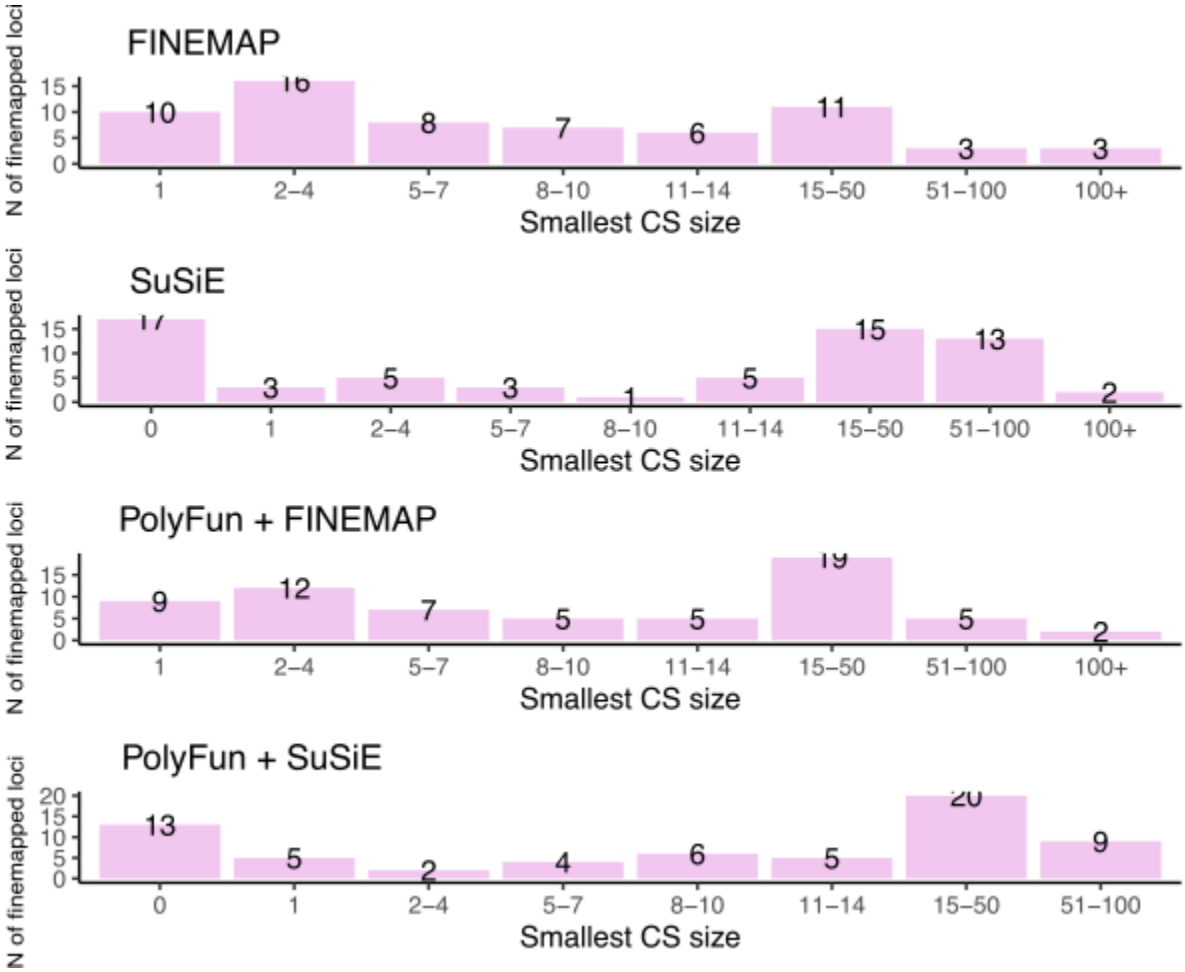


Figure S3 A-C Distribution of fine-mapped loci according to the smallest 95% credible set (95% CS) formed, using different fine-mapping methods and LD reference panels. Fine-mapping methods include: FINEMAP, SuSiE, PolyFun + FINEMAP, PolyFun + SuSiE. 'Smallest CS size' denotes the number of SNPs comprising the smallest 95% CS for a given fine-mapped locus. Absolute numbers denote the number of fine-mapped loci in each 'smallest CS size' category. If the 'smallest CS size' is 0, this denotes that no 95% CS was formed. Note: 95% CS for FINEMAP is a set of SNPs of which the joint posterior probability of including the causal SNP(s) is higher than 0.95. 95% CS for SuSiE denotes the sum of the PIPs equals to 95%, in which case each PIP is a marginal posterior probability for a single SNP. HRC - Haplotype Reference Consortium, UKB - UK Biobank.

A). HRC LD reference panel



B). UKB LD reference panel



C). In-sample LD

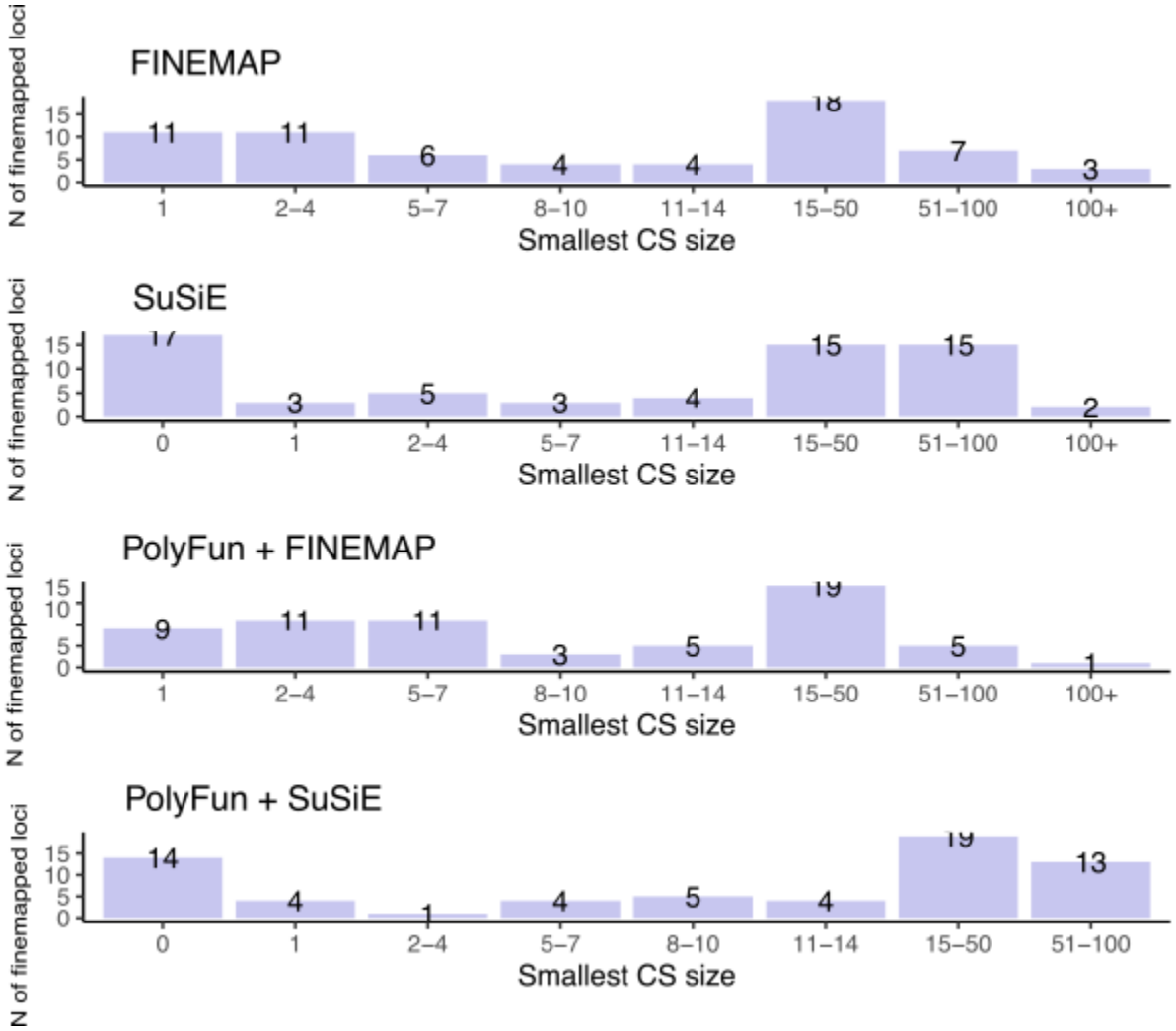


Figure S4. Plot of consensus SNPs with PIP >0.50 across all 16 fine-mapping analyses, including different LD reference panels and fine-mapping methods. The color of the points corresponds to the LD information used: UK Biobank (pink), Haplotype Reference Consortium (blue), in-sample LD (purple) and no LD (single variant fine-mapping) (grey). Small shapes denote SNPs with PIP between 0.50 and 0.95, while the large shapes denote SNPs with PIP above 0.95. Circles indicate statistical fine-mapping methods and squares indicate functional fine-mapping methods. On the y axis, the PGC3 locus name is displayed in parenthesis after each fine-mapped SNP and indicates the name assigned to identify the locus in the original PGC3 GWAS publication, which is not necessarily the causal gene.

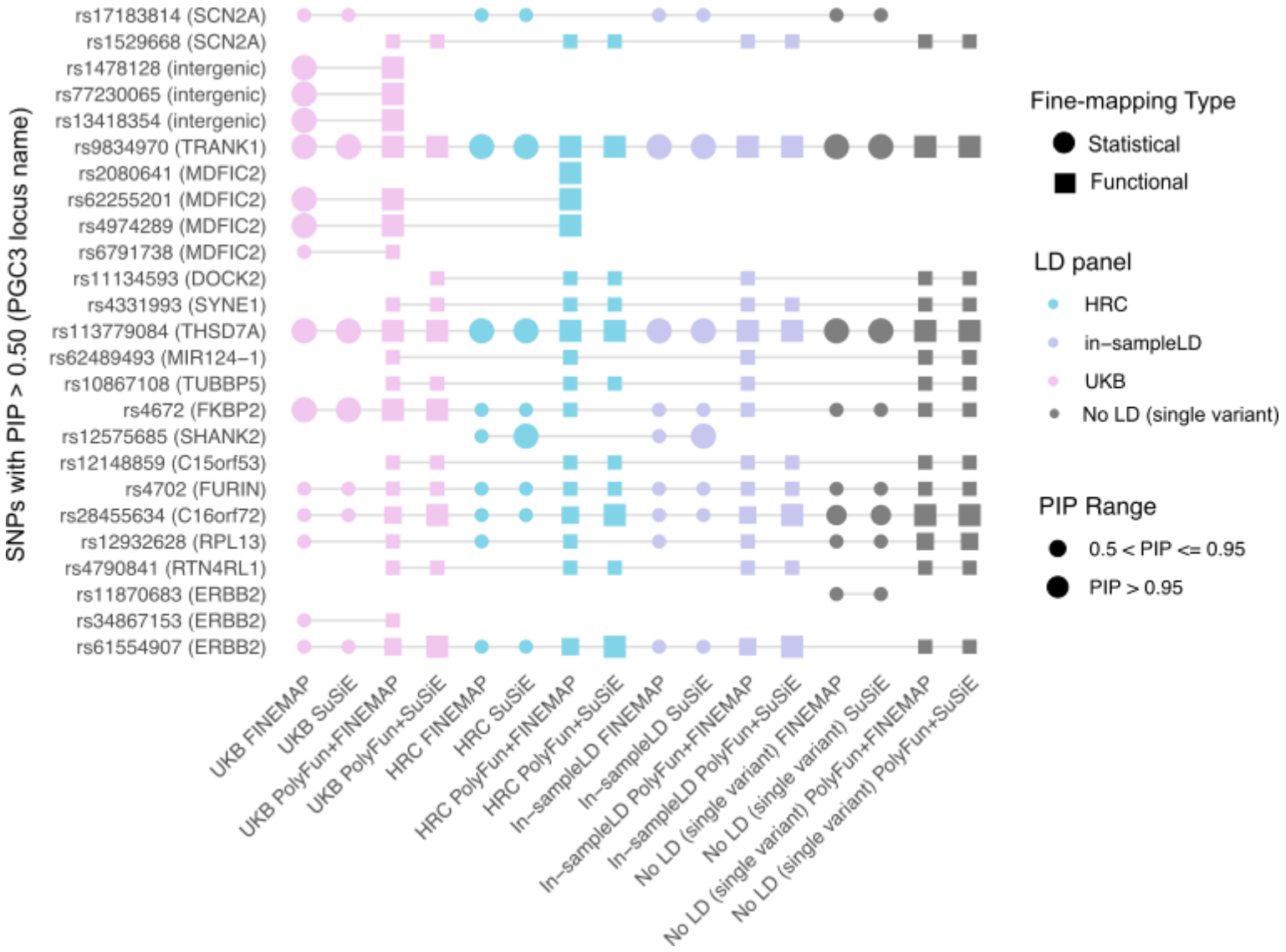
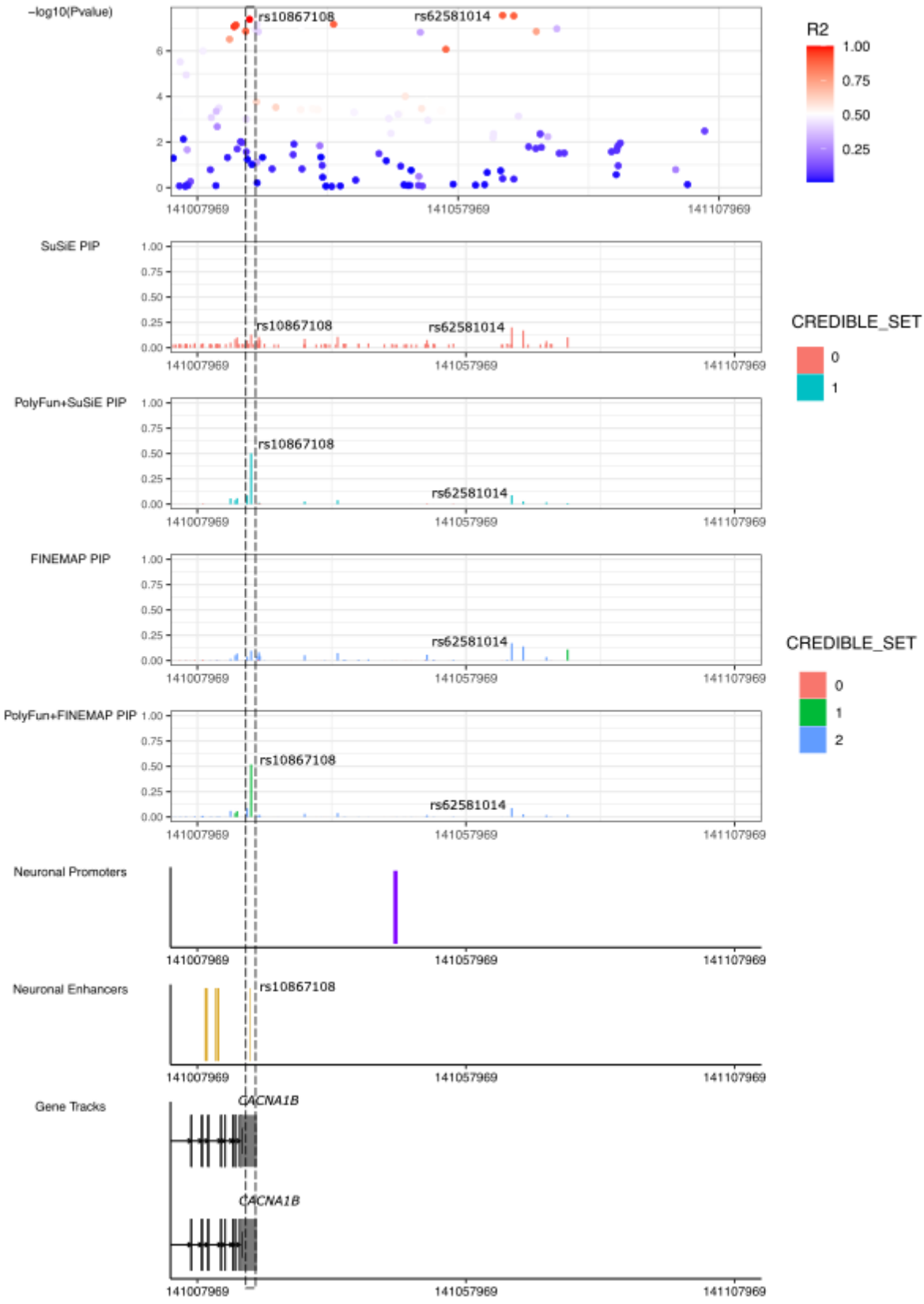
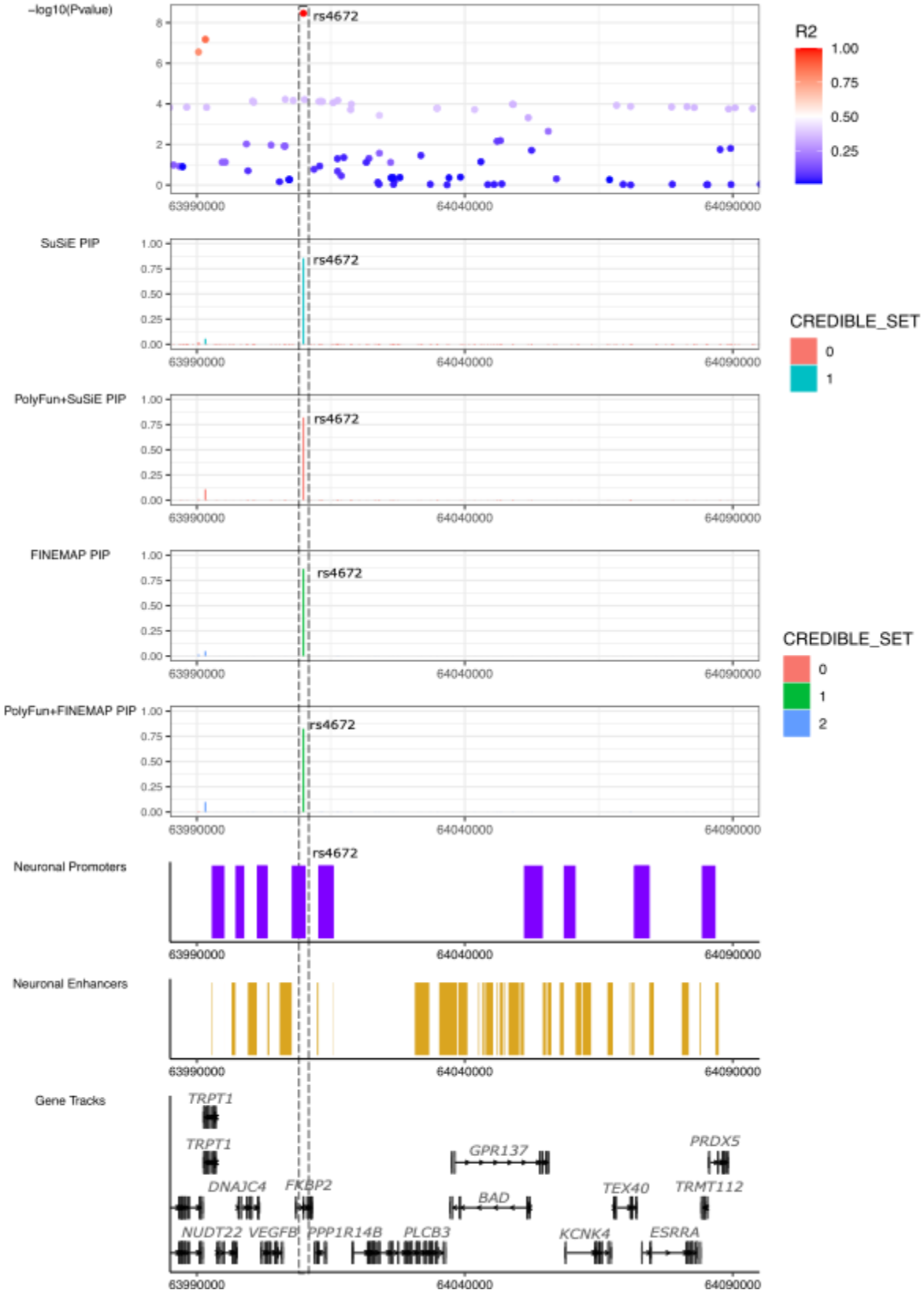


Figure S5 A-N. Multi-track locus plots for fine-mapped loci according to convergence of evidence across validation analyses. Panels A-N are named according to the genome-wide significant (GWS) locus name assigned to identify the locus in the PGC3 BD GWAS (Mullins et al., 2021) (which is not necessarily the causal gene). The upper track depicts the GWAS association statistics over a window of 100,000 - 200,000 base pairs. Variants are colored according to their linkage disequilibrium (LD) r^2 with the index SNP, calculated based on the HRC reference panel. In the following four tracks, posterior inclusion probabilities (PIPs) are provided from SuSiE, PolyFun+SuSiE, FINEMAP and PolyFun+FINEMAP based on the Haplotype Reference Consortium LD reference panel. Each bar represents a variant in a credible set, and the bars are colored according to the credible set they belong to. The lower tracks visualize overlap with neuronal enhancers or promoters (Nott et al., 2019) and the gene tracks (Ensembl v75). On each track, the SNPs labeled represent the SNP prioritized through fine-mapping and the index SNP in the GWS locus. All genomic coordinates are in GRCh37.

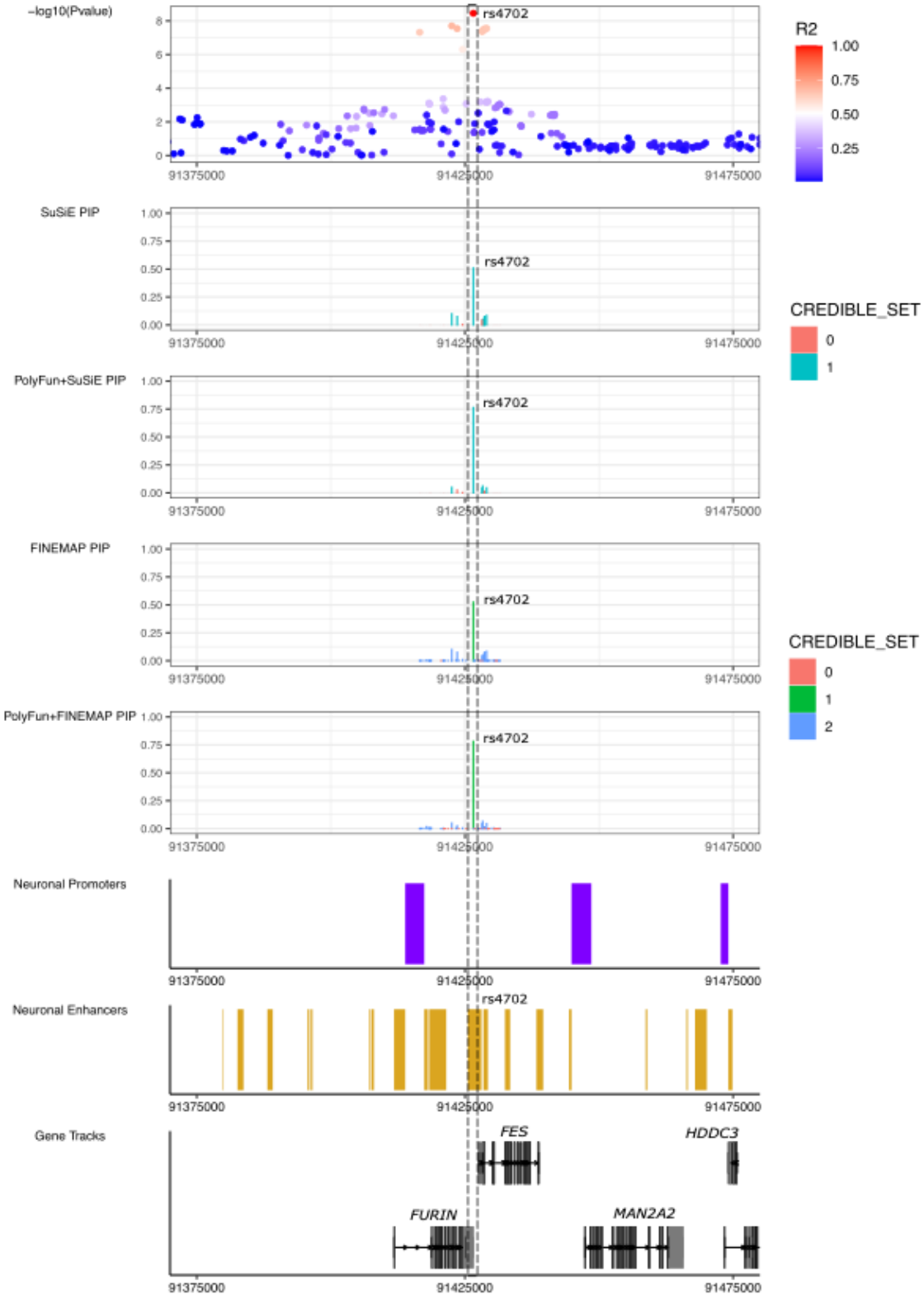
A). **TUBBP5 locus** (Chromosome 9) (SNP with PIP> 0.50 and part of a 95% CS, SMR eQTL, sQTL and mQTL evidence). Coordinates of locus fine-mapped = chr9: 141,007,969 - 141,110,969.



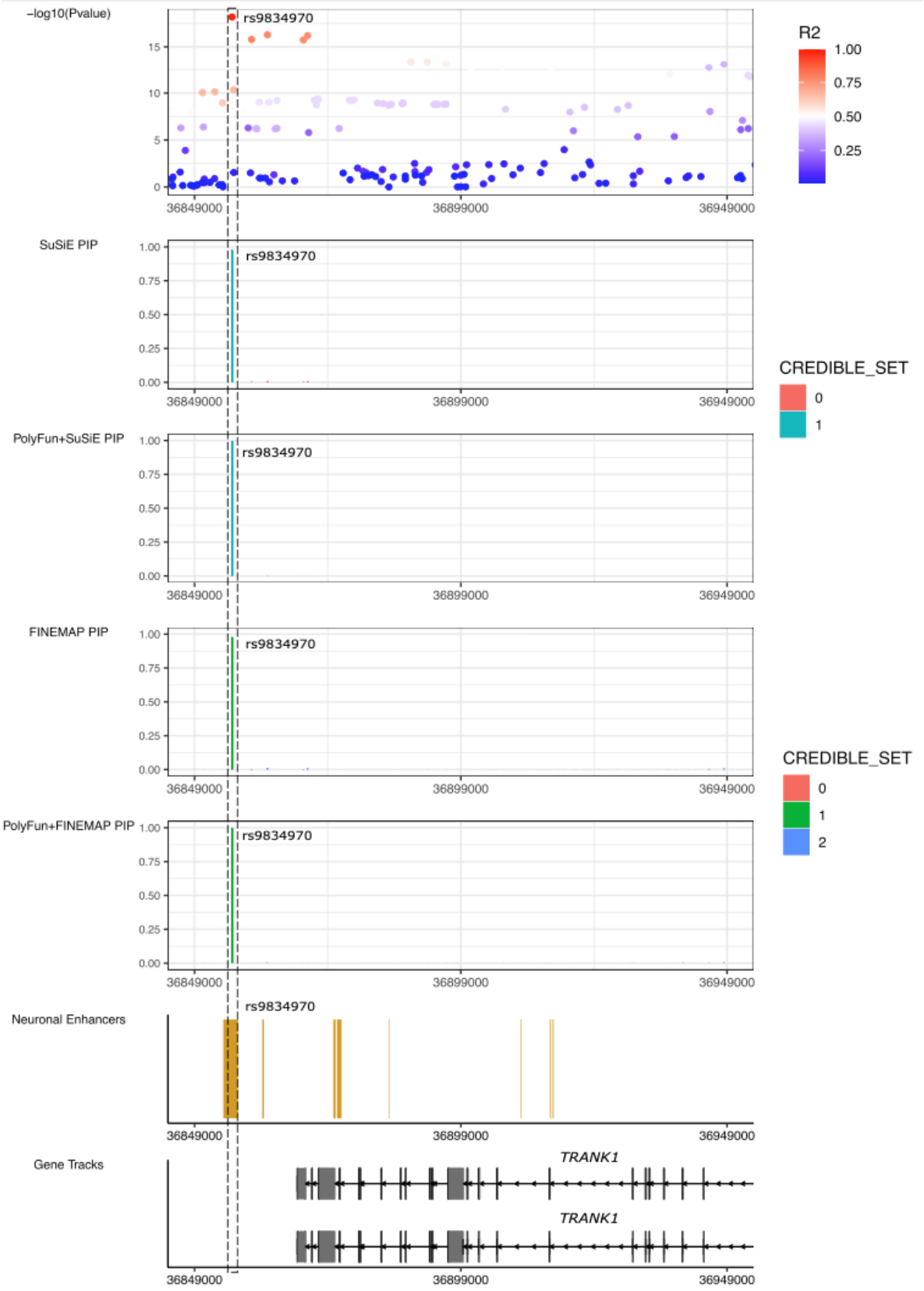
B). **FKBP2 locus** (Chromosome 11) (SNP with PIP> 0.50 and part of a 95%CS, missense variant, SMR eQTL, sQTL and mQTL evidence, overlaps within astrocyte and neuronal promoters). Coordinates of locus fine-mapped = chr11: 63,990,000 - 64,090,000.



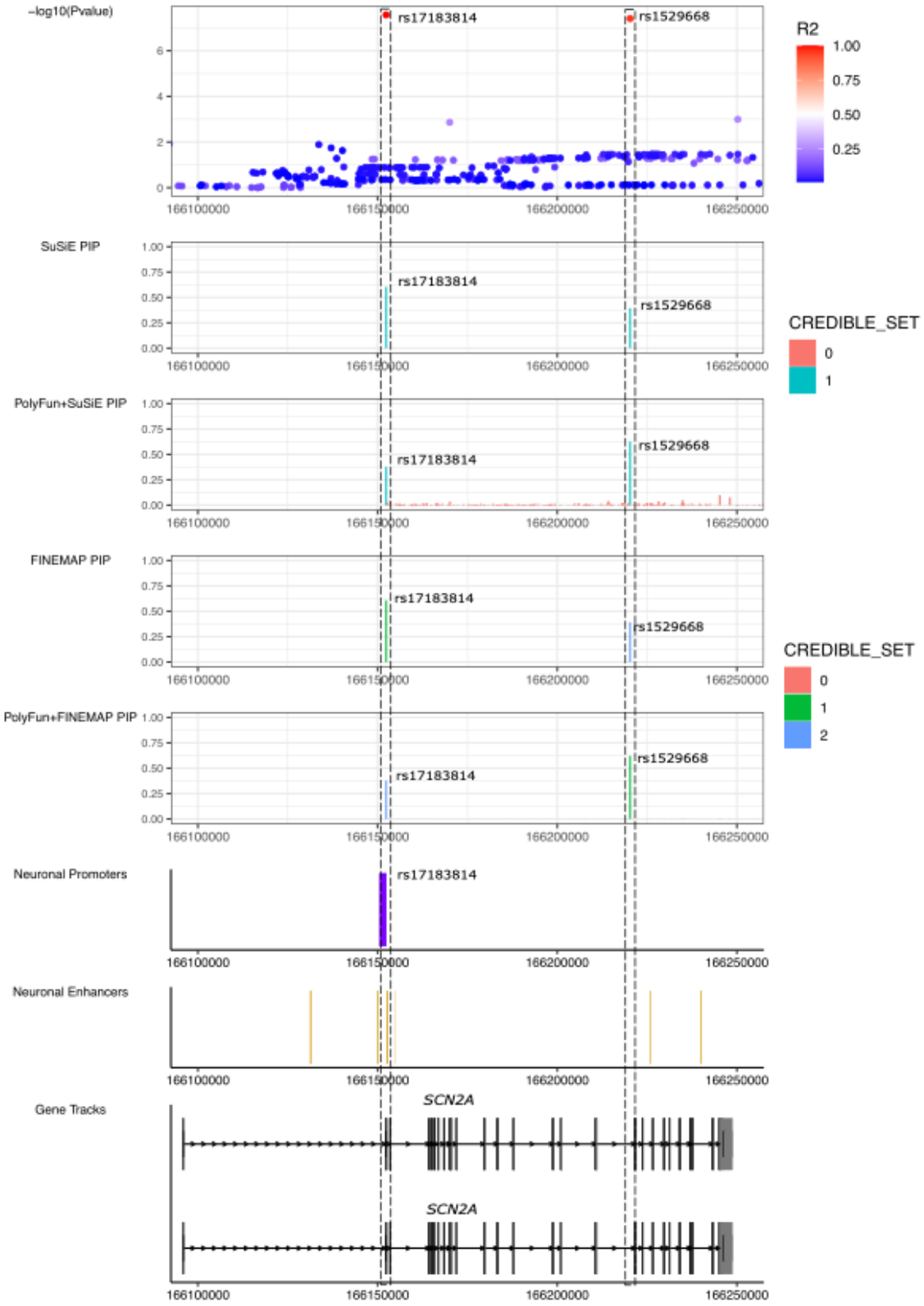
C). **FURIN locus** (Chromosome 15) (SNP with PIP> 0.50 and part of a 95%CS, SMR eQTL, sQTL and mQTL evidence, enhancer-promoter interaction through PLAC-seq). Coordinates of locus fine-mapped = chr15: 91,375,000 - 91,475,000.



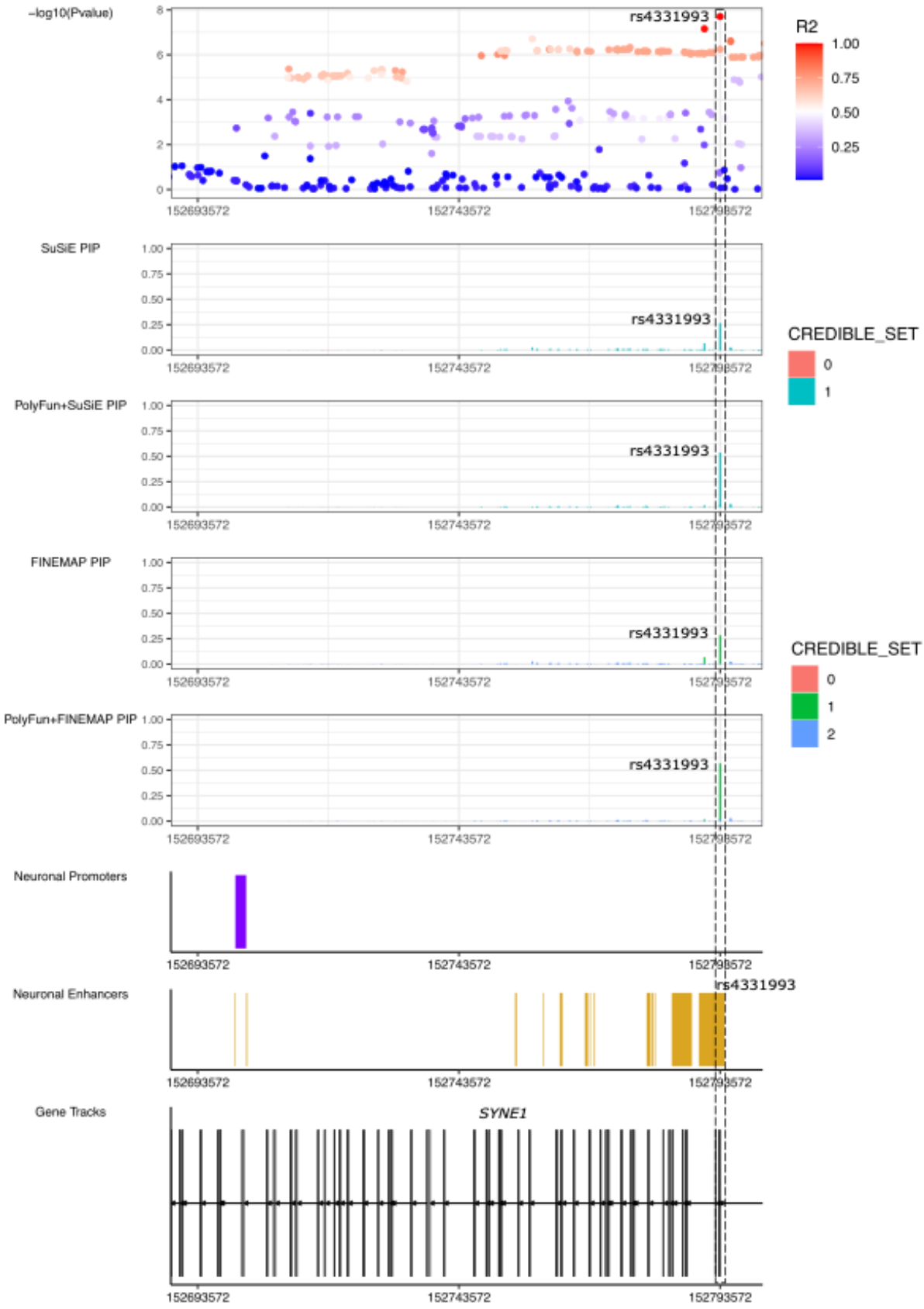
D). **TRANK1 locus** (Chromosome 3) (SNP with PIP> 0.50 and part of a 95%CS, SMR eQTL evidence, overlap within neuronal enhancer). Coordinates of locus fine-mapped = chr3: 36,849,000 - 36,949,000.



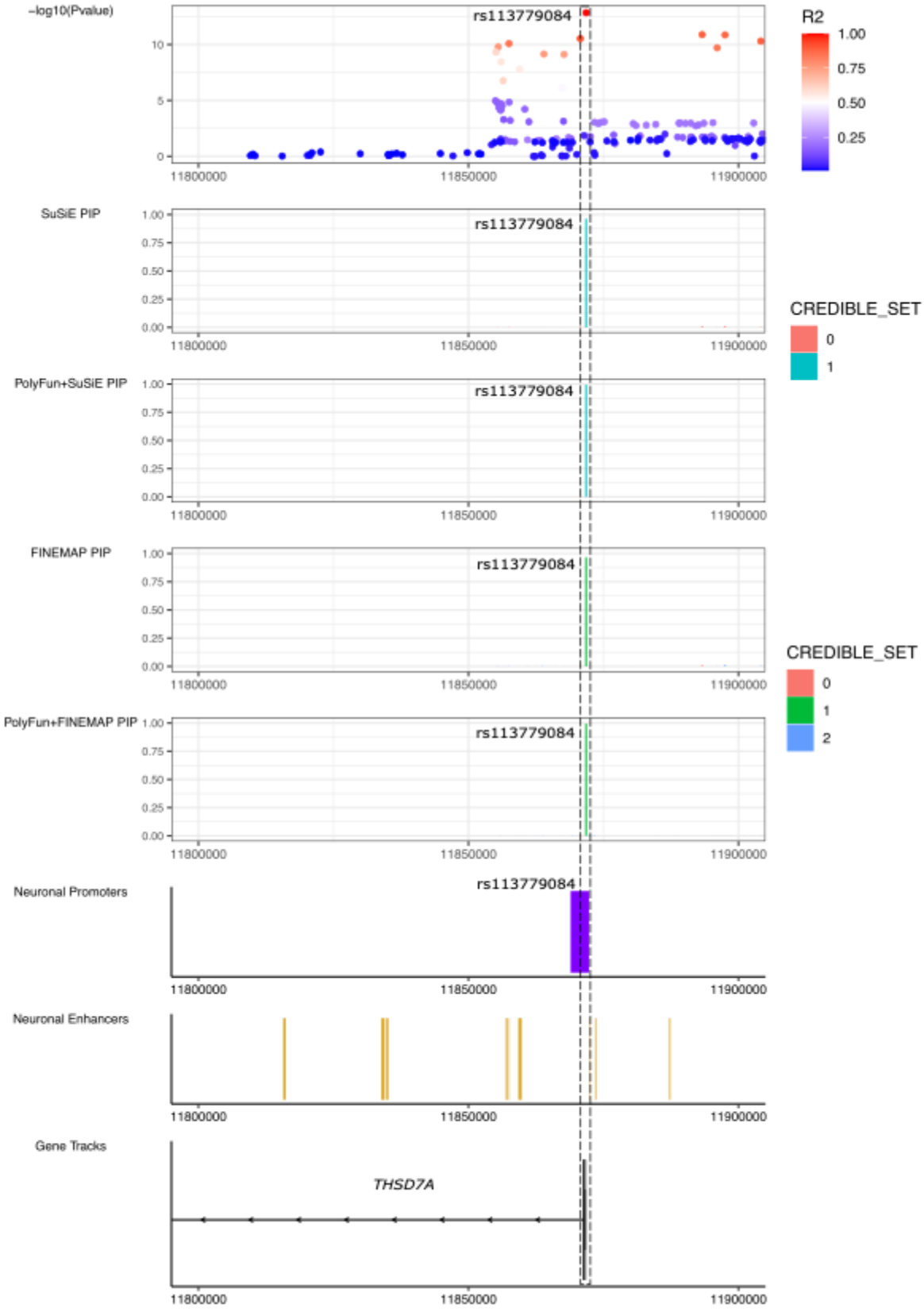
E). **SCN2A locus** (Chromosome 2)(SNP with PIP> 0.50 and part of a 95%CS, missense variant, SMR sQTL evidence, overlap within neuronal promoter). Coordinates of locus fine-mapped = chr2: 166,100,000 - 166,250,000.



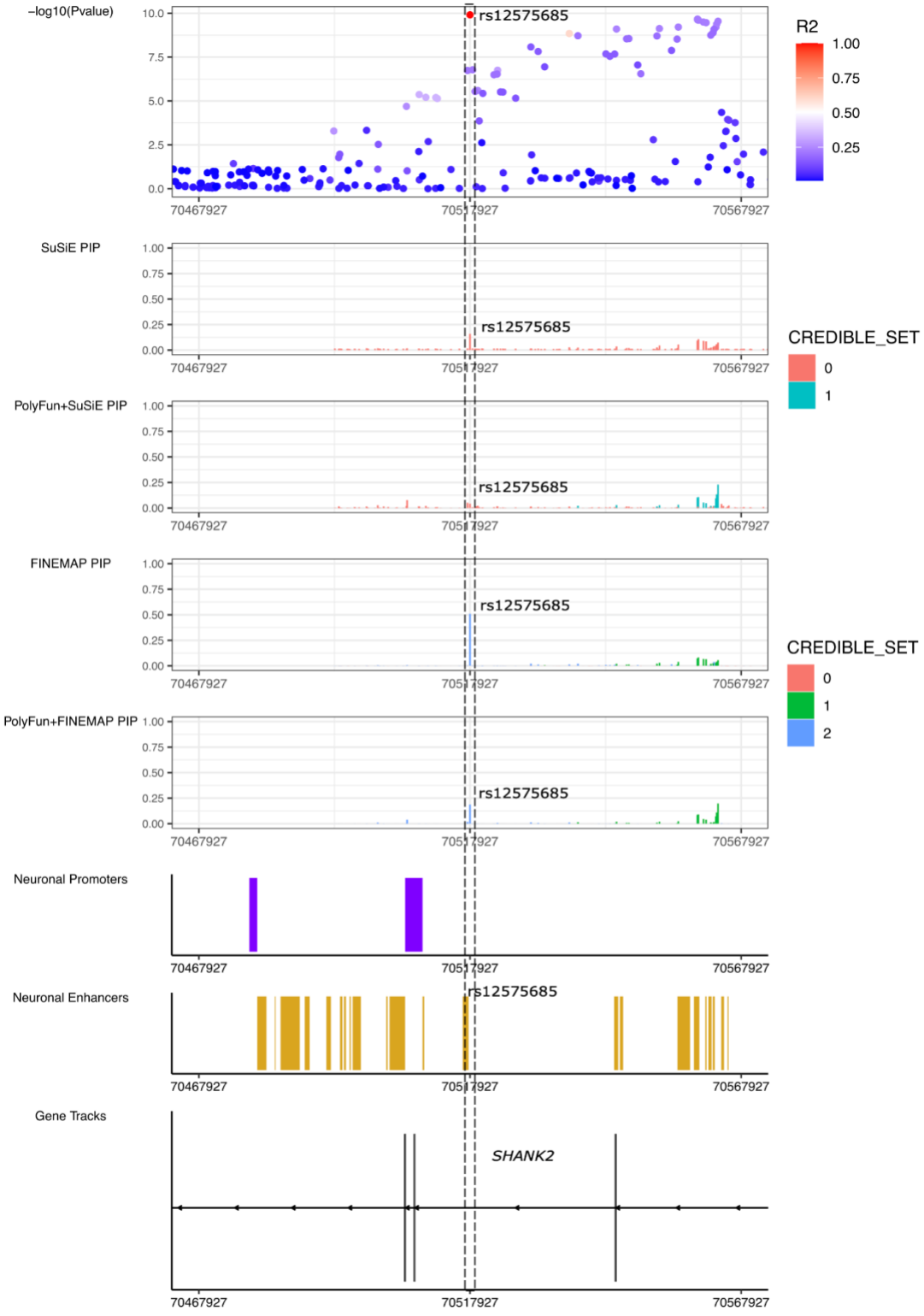
F). **SYNE1 locus** (Chromosome 6) (SNP with PIP> 0.50 and part of a 95%CS, enhancer-promoter interaction through PLAC-seq). Coordinates of locus fine-mapped = chr6: 152,693,572 - 152,796,572.



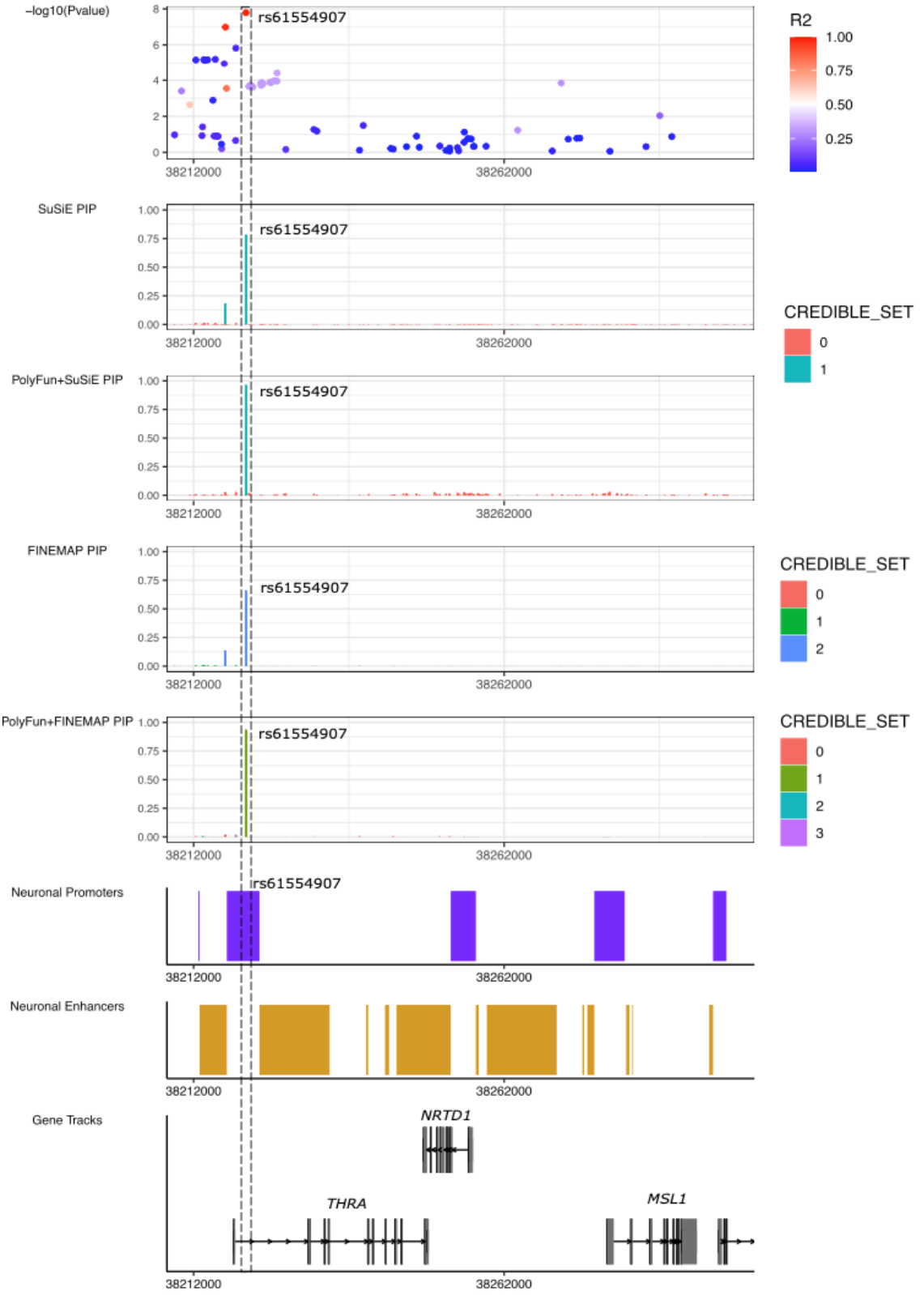
G). **THSD7A locus** (Chromosome 7) (SNP with PIP> 0.50 and part of a 95%CS, overlap within astrocyte and neuronal promoters). Coordinates of locus fine-mapped = chr7: 11,800,000 - 11,900,000.



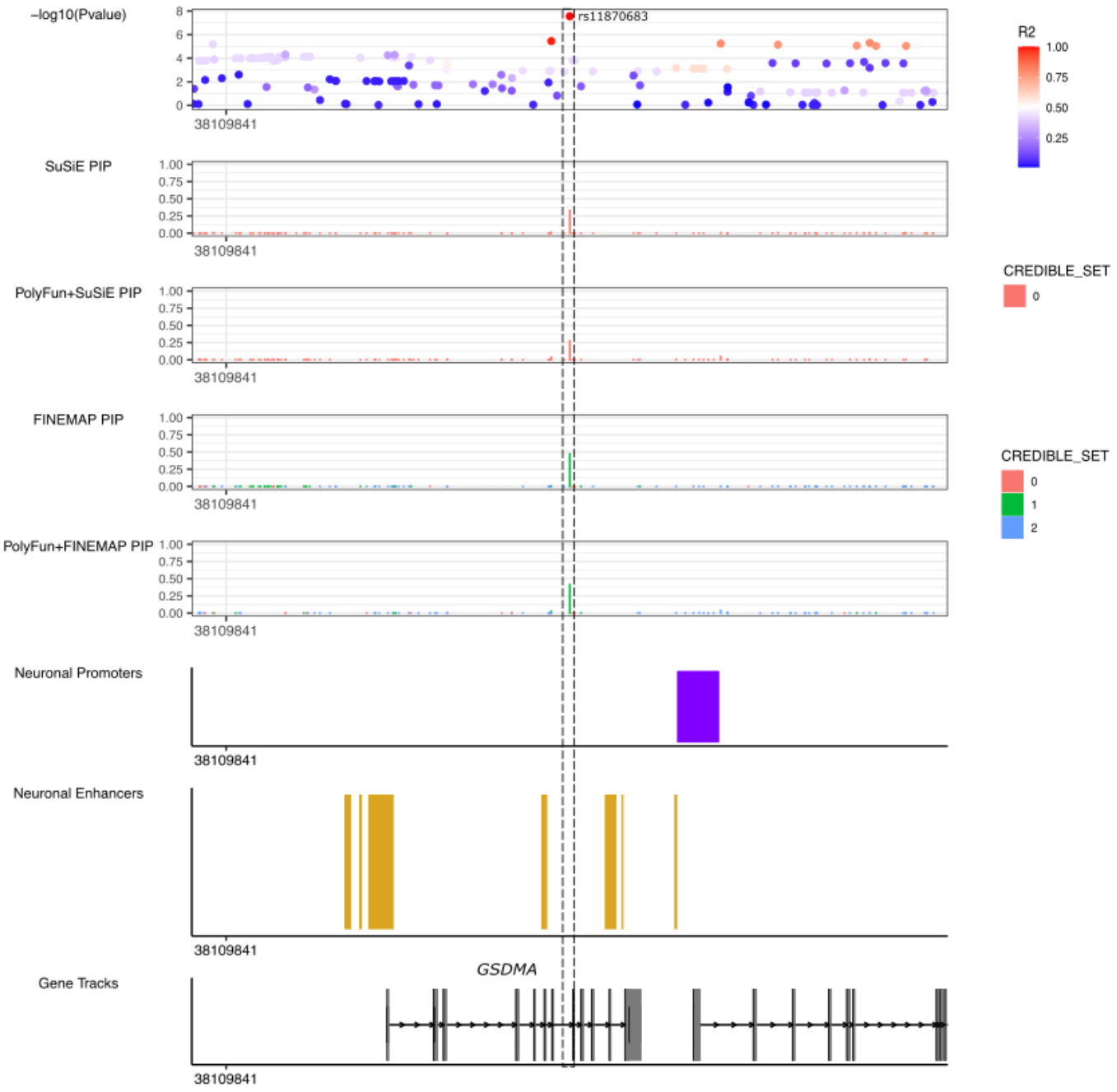
H). **SHANK2 locus** (Chromosome 11; rs12575685) (SNP with PIP> 0.50 and part of a 95%CS). Coordinates of locus fine-mapped = chr11: 70,467,927 - 70,567,927.



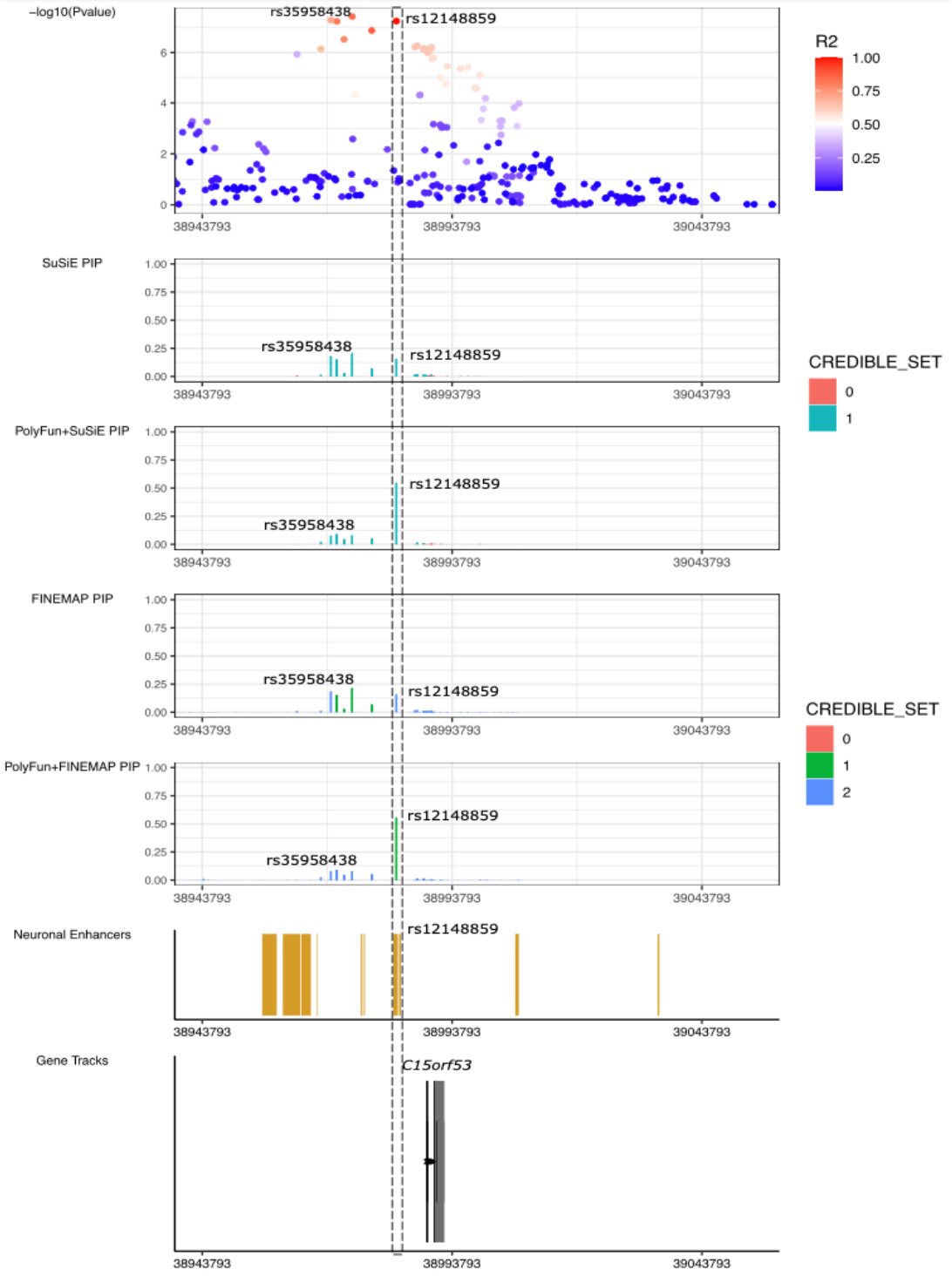
I).**ERBB2 locus** (Chromosome 17; rs61554907) (SNP with PIP> 0.50 and part of a 95%CS, overlap within astrocyte and neuronal promoters). Coordinates of locus fine-mapped = chr17: 38,212,000 - 38,298,000.



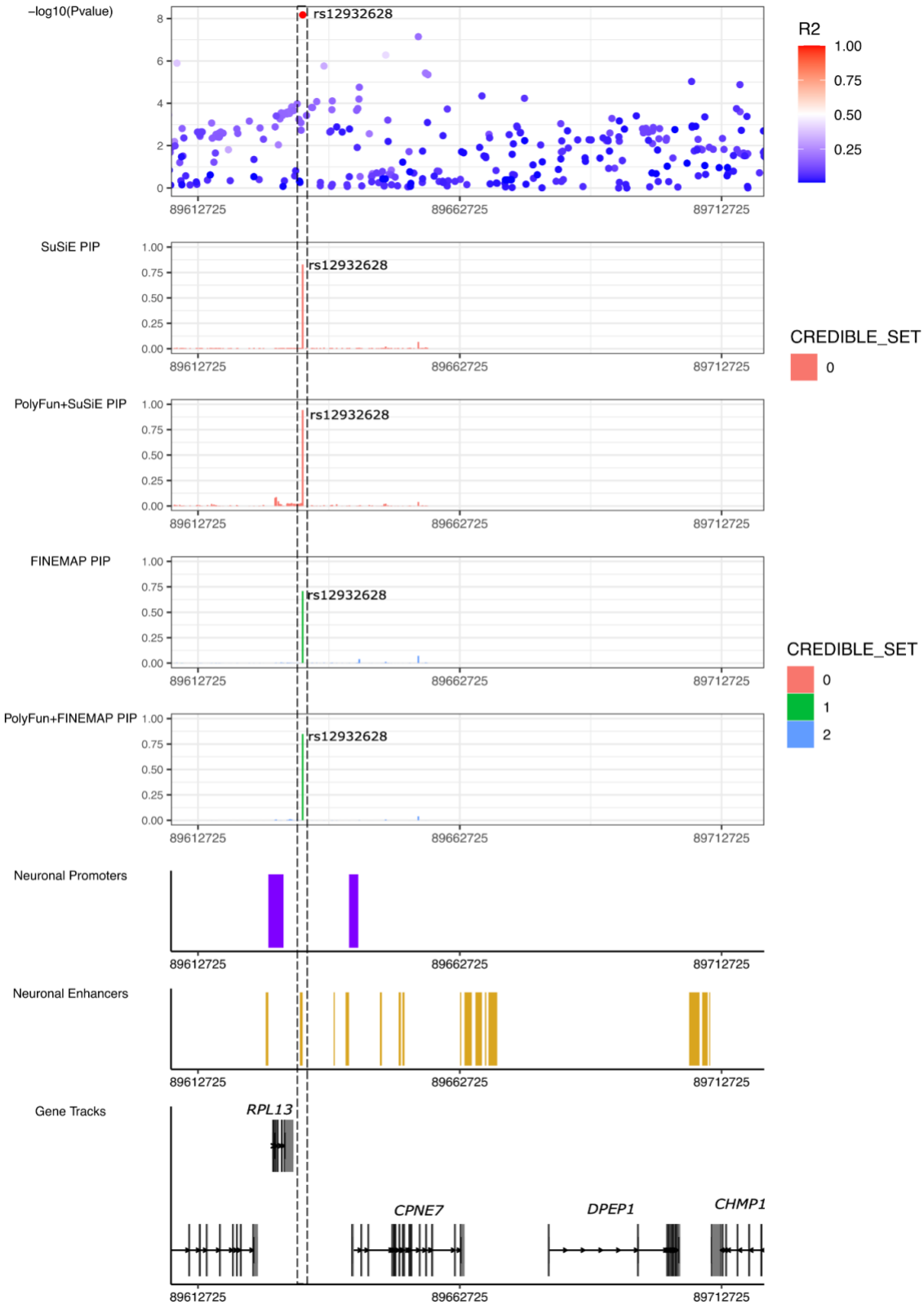
J).**ERBB2 locus** (Chromosome 17; rs11870683) (SNP with PIP> 0.50 and part of a 95%CS, SMR sQTL and mQTL evidence). Coordinates of locus fine-mapped = chr17: 38,109,841 - 38,149,841.



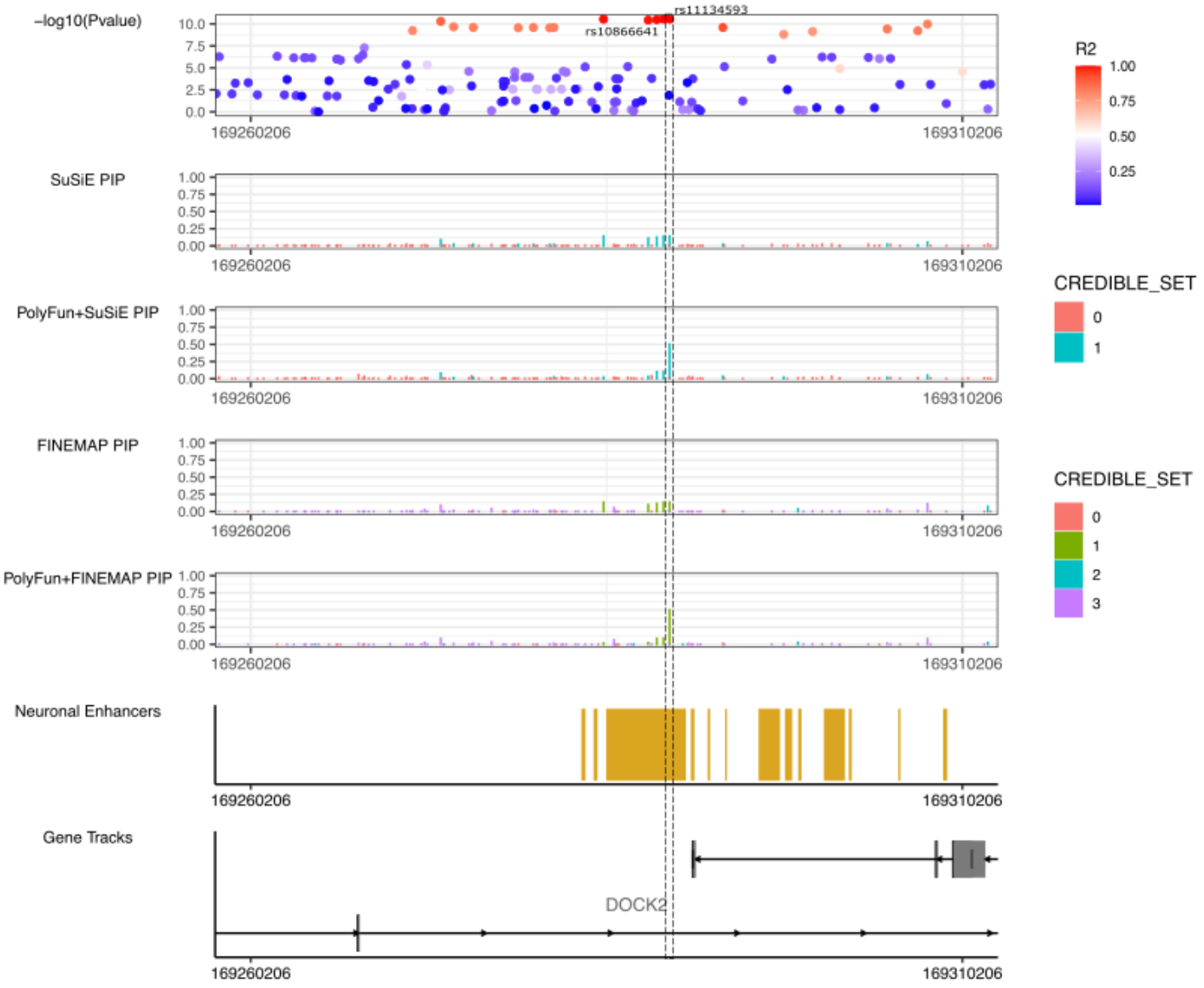
K).**C15orf53 locus** (Chromosome 15) (SNP with PIP> 0.50 and part of a 95%CS, enhancer-promoter interaction through PLAC-seq, overlap within neuronal enhancers). Coordinates of locus fine-mapped = chr15: 38,943,793 - 39,053,793.



L).**RPL13 locus** (Chromosome 16) (SNP with PIP> 0.50 and part of a 95%CS). Coordinates of locus fine-mapped = chr16: 89,612,725 - 89,715,725.



M)**. DOCK2 locus (Chromosome 5)** (SNP with PIP> 0.50 and part of a 95%CS, enhancer-promoter interaction through PLAC-seq, overlap within neuronal enhancers). Coordinates of locus fine-mapped = chr5: 169,260,206 - 169,310,206.



N)**. C16orf72 locus (Chromosome 16)** (SNP with PIP> 0.50 and part of a 95%CS). Coordinates of locus fine-mapped = chr16: 9,130,816 - 9,330,816.

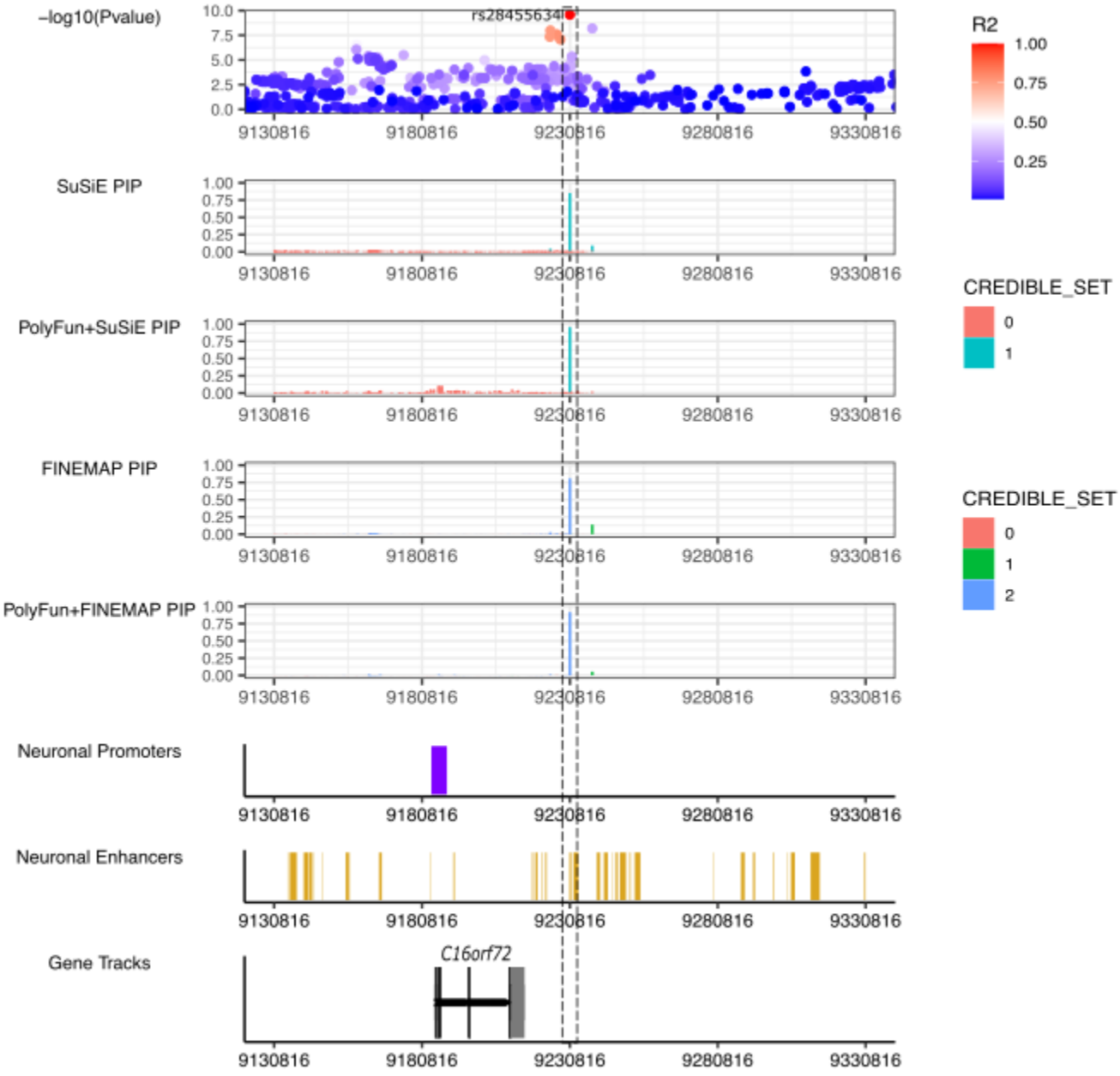
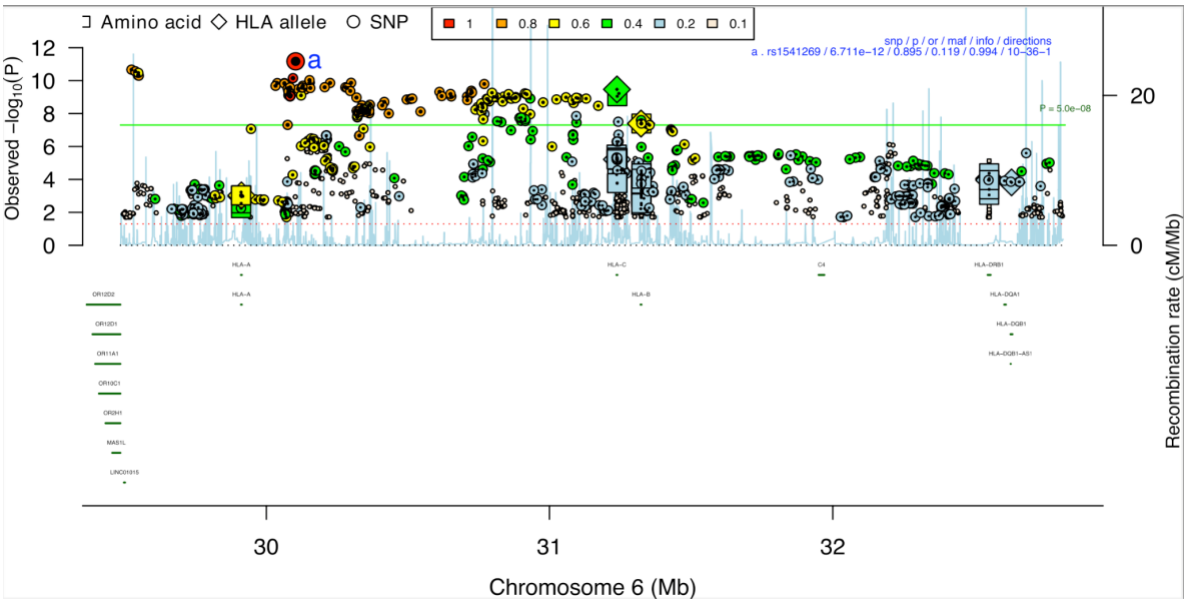
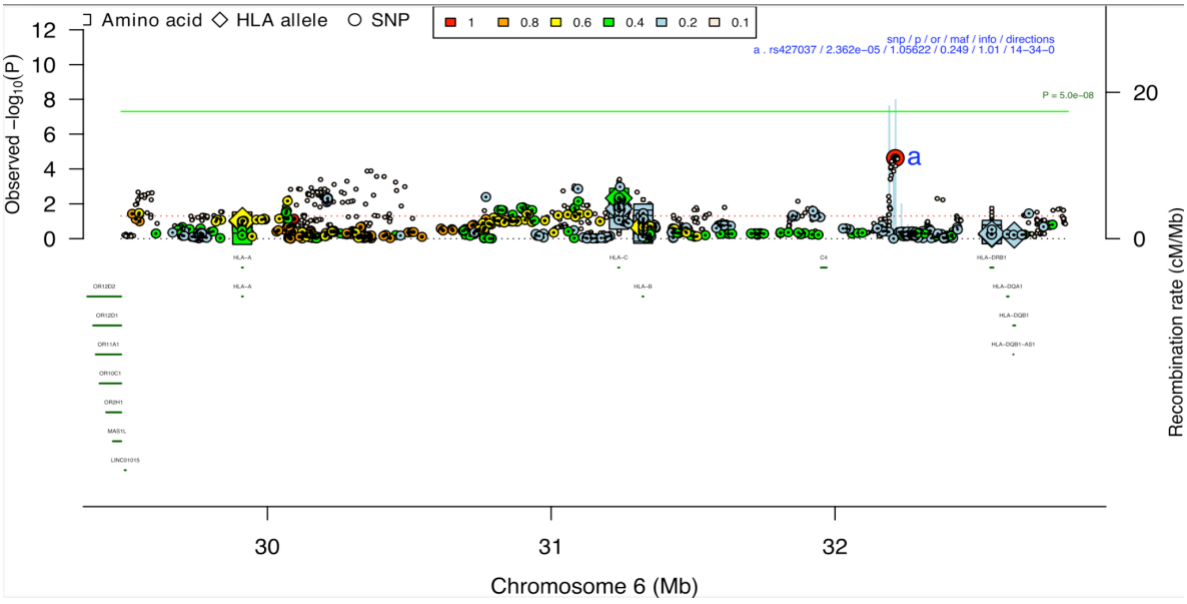


Figure S6. Area plots of the MHC locus after imputation of HLA variants and amino acids based on the 1000 genomes Phase 3 (European ancestry) reference panel. Panel A shows the association results from the meta-analysis of 48 BD cohorts before conditioning on the top lead SNP, while panel B shows the association results after conditioning on the top lead SNP (rs1541269). The color of the variants corresponds to their linkage disequilibrium r^2 value with the index variant in each panel. The shape of each point corresponds to the type of variant - SNP (circle), amino acid (square) or HLA allele (diamond).

A).



B).



Supplementary Note

Comparison of the impact of LD reference panels on fine-mapping

We aimed to investigate the impact of using an LD reference panel during fine-mapping, compared with using LD information calculated from the original GWAS data. The latter is typically considered the gold-standard approach, however is difficult in practice due to data availability and sharing restrictions.

We had access to the individual-level genetic data for 48/57 cohorts in the PGC BD GWAS, comprising 33,781 BD cases and 53,869 controls. First, we performed a meta-analysis of the GWAS summary statistics for these 48 cohorts ("meta-48") using the same procedures as the original GWAS[1], followed by cleaning of the GWAS results using DENTIST software. These "meta-48" results comprised 7,616,190 SNPs, and there were 45 GWS loci. We calculated "in-sample" LD using the individual-level genetic data from these 48 cohorts, as described in the main text Methods section on "LD reference panels". We fine-mapped 44 GWS loci (excluding the MHC locus), using the GWAS summary statistics from the "meta-48", using 4 fine-mapping methods, the GWS locus window and a 3Mb window, using "in-sample" LD calculated based on the exact same set of individuals in the "meta-48" GWAS. These analyses were repeated with the three other LD options: 1) the HRC reference panel, 2) the UKB reference panel and 3) single variant fine-mapping with no LD. The results of each of these fine-mapping analyses were compared against the corresponding analysis based on the "in-sample" LD, using the Jaccard Index (or Jaccard Similarity Coefficient), to summarize the concordance in the results between them. The Jaccard Index was calculated as the number of fine-mapped SNPs (PIP > 0.5 and in a 95% CS) in both fine-mapping methods (intersection), divided by the total number of fine-mapped SNPs across either method (union), and ranges from 0 (no concordance between the methods) to 1 (complete concordance between the methods). We computed the Jaccard Index across the different LD panels and windows and compared them against those derived from in-sample LD. We observed high levels of concordance between the fine-mapping results based on in-sample LD compared to the HRC (mean Jaccard Index 0.94 (sd 0.12)), followed by UKB (mean Jaccard Index 0.84 (sd 0.10)), and a lesser degree of concordance with the no LD (single variant) fine-mapping (mean Jaccard Index 0.67 (sd 0.05)) (**Supplementary Note Table 1**).

Supplementary Note Table 1: Jaccard indices comparing fine-mapping results based on different LD options with those based on "in-sample" LD. Fine-mapping was performed on the "meta48 GWAS". Analyses include 4 fine-mapping methods, 2 fine-mapping windows and 3 LD options compared against in-sample LD. Results are based on fine-mapped SNPs with PIP > 0.50 and parts of 95% credible sets. Brackets after each Jaccard index denote the number of SNPs featuring in both sets (intersection), divided by the unique number of SNPs across both sets (union).

Fine-mapping window_method	HRC	UKB	No LD (single variant)
GWS locus FINEMAP	1 (7/7)	0.88 (7/8)	0.71 (5/7)
GWS locus SuSiE	1 (7/7)	1 (7/7)	0.71 (5/7)
GWS locus PolyFun+FINEMAP	1 (8/8)	0.88 (8/9)	0.67 (6/9)
GWS locus PolyFun+SuSiE	1 (6/6)	0.86 (6/7)	0.63 (5/8)

3Mb FINEMAP	0.71 (5/7)	0.66 (4/6)	0.6 (3/5)
3Mb SuSiE	1 (4/4)	0.8 (4/5)	0.75 (3/4)
3Mb PolyFun+FINEMAP	0.78 (7/9)	0.875 (7/8)	0.625 (5/8)
3Mb PolyFun+SuSiE	1 (4/4)	0.75 (3/4)	0.67 (4/6)
Mean (sd)	**0.94 (0.12)**	**0.84 (0.10)**	**0.67 (0.05)**

Second, we calculated the correlation of the PIP values of SNPs fine-mapped (PIP > 0.5 and parts of 95% credible sets) using both the in-sample LD and the specified LD option (HRC, UKB and no LD (single variant fine-mapping) (**Supplemental Note Table 2**). Amongst SNPs fine-mapped by both in-sample LD and each specified LD option, PIP values were extremely highly correlated.

Supplementary Note Table 2: Correlation (r) of posterior inclusion probability (PIP) values of SNPs fine-mapped using both the in-sample LD and the specified LD option. Fine-mapping was performed on the "meta48 GWAS". Analyses include 4 fine-mapping methods, 2 fine-mapping windows and 3 LD options compared against in-sample LD. Results are based on fine-mapped SNPs with PIP > 0.50 and parts of 95% credible sets. Values in brackets after each correlation denote the number of SNPs fine-mapped using both the specified LD option and in-sample LD.

Fine-mapping window_method	**HRC**	**UKB**	**No LD (single variant)**
GWS locus FINEMAP	0.998 (7)	0.999 (7)	0.970 (5)
GWS locus SuSiE	0.999 (7)	0.999 (7)	0.987 (5)
GWS locus PolyFun+FINEMAP	0.994 (8)	0.989 (8)	0.998 (6)
GWS locus PolyFun+SuSiE	0.999 (6)	0.999 (6)	0.999 (5)
3Mb FINEMAP	0.999 (5)	0.999 (4)	0.999 (3)
3Mb SuSiE	0.999 (4)	0.997 (4)	0.999 (3)
3Mb PolyFun+FINEMAP	0.999 (7)	0.999 (7)	0.920 (5)
3Mb PolyFun+SuSiE	0.999 (4)	0.997 (3)	0.902 (4)
Mean (sd)	**0.998 (0.002)**	**0.997 (0.003)**	**0.972 (0.039)**

Third, we assessed the correlation of the allele frequencies calculated from the "in-sample" LD dataset and the HRC reference panel. Between the HRC and in-sample LD (7,611,652 SNPs from the "meta48" GWAS summary statistics found in both), the correlation of allele frequencies was 0.99, further emphasizing the comparability of the two datasets. Allele frequency calculations were not possible for

the UK Biobank LD reference panel used, as data were provided as LD matrices rather than individual-level genetic data.

In summary, analyses of the impact of LD reference panels on fine-mapping demonstrate the strong concordance of results between the HRC reference panel and the in-sample LD, supporting our hypothesis that a well-matched LD reference panel to the GWAS summary statistics can achieve high fine-mapping resolution, when in-sample LD estimates are not available.

Comparison of the impact of locus windows on fine-mapping

To assess the impact of locus windows on fine-mapping, we fine-mapped the 44 GWS loci from the meta48 GWAS using the GWS locus window (defined as in the original GWAS as LD $r^2 > 0.1$ with the top lead SNP within a 3 Mb range) and a 3Mb fine-mapping window. The 3 Mb fine-mapping windows were selected using the 2,763 published LD matrices of 3 Mb blocks covering the entire genome calculated in the UKB, such that the index SNP was as close as possible to the center of the 3Mb window. We used the coordinates of these 3 Mb windows, as they are widely used to obtain LD information for post-GWAS analyses, are provided for download with the Polyfun and PolyPred softwares, and represent a large (n=337,000) publicly available European ancestry LD reference panel. Keeping the LD option and fine-mapping method consistent, we calculated the Jaccard Indices of fine-mapped SNPs (defined as SNPs with PIP > 0.5 and in 95% credible sets) between the GWS locus and 3Mb windows (**Supplemental Note Table 3).** Comparing the two fine-mapping windows, Jaccard Indices ranged from 0.43 to 0.86 across analyses, with mean 0.59 (sd 0.12).

Supplementary Note Table 3: Jaccard indices comparing fine-mapping results from GWS locus versus 3Mb fine-mapping windows. Fine-mapping was performed on the "meta48 GWAS". Analyses include 4 fine-mapping methods and 4 LD options and 2 fine-mapping windows were compared. Results are based on fine-mapped SNPs with PIP > 0.50 and parts of 95% credible sets. Brackets after each Jaccard Index denote the number of SNPs featuring in both sets (intersection), divided by the unique number of SNPs across both sets (union).

LD option_Fine-mapping method	**GWS locus vs. 3Mb fine-mapping window**
In-sample LD FINEMAP	0.5 (4/8)
In-sample LD SuSiE	0.57 (4/7)
In-sample LD PolyFun+FINEMAP	0.5 (5/10)
In-sample LD Polyfun+SuSiE	0.5 (4/8)
HRC FINEMAP	0.55 (5/9)
HRC SuSiE	0.57 (4/7)

HRC PolyFun+FINEMAP	0.54 (6/11)
HRC Polyfun+SuSiE	0.5 (4/8)
UKB FINEMAP	0.625 (5/8)
UKB SuSiE	0.714 (5/7)
UKB PolyFun+FINEMAP	0.55 (6/11)
UKB Polyfun+SuSiE	0.43 (3/7)
No LD (single variant) FINEMAP	0.6 (3/5)
No LD (single variant) SuSiE	0.6 (3/5)
No LD (single variant) PolyFun+FINEMAP	0.857 (6/7)
No LD (single variant) Polyfun+SuSiE	0.857 (6/7)
Mean (sd)	**0.59 (0.12)**

We calculated the correlation of the PIP values of SNPs fine-mapped (PIP > 0.5 and parts of 95% credible sets) using both the GWS locus and 3 Mb fine-mapping window (**Supplemental Note Table 4**). Despite the modest Jaccard Indices of concordance in fine-mapping results between the two fine-mapping windows, the PIP correlation values between SNPs successfully fine-mapped using both fine-mapping windows were high (mean PIP correlation 0.988 (sd 0.025)).

Supplementary Note Table 4: Correlation (r) of posterior inclusion probability (PIP) values of fine-mapped SNPs identified in both fine-mapping windows (GWS locus vs 3Mb). These results are based on GWAS meta-analysis using only the 48 cohorts for which we have individual level genetic data available. The analyses include 4 LD options and 4 fine-mapping methods, and they are based on fine-mapped SNPs with PIP > 0.50 and parts of 95% credible sets. Values in the brackets denote the number of shared fine-mapped SNPs amongst the different LD panels and in-sample LD.

LD option_Fine-mapping method	**GWS locus vs. 3Mb fine-mapping window**
In-sample LD FINEMAP	0.997 (4)
In-sample LD SuSiE	0.997 (4)
In-sample LD PolyFun+FINEMAP	0.999 (5)

In-sample LD Polyfun+SuSiE	0.999 (4)
HRC FINEMAP	0.997 (5)
HRC SuSiE	0.998 (4)
HRC PolyFun+FINEMAP	0.995 (6)
HRC Polyfun+SuSiE	0.999 (4)
UKB FINEMAP	0.997 (5)
UKB SuSiE	0.997 (5)
UKB PolyFun+FINEMAP	0.995 (6)
UKB Polyfun+SuSiE	0.997 (3)
No LD (single variant) FINEMAP	0.999 (3)
No LD (single variant) SuSiE	0.999 (3)
No LD (single variant) PolyFun+FINEMAP	0.926 (6)
No LD (single variant) Polyfun+SuSiE	0.921 (6)
Mean (sd)	**0.988 (0.025)**

In summary, our comparison of the impact of LD reference panels and fine-mapping windows on the "meta48" GWAS, indicated that fine-mapping results were more similar when using different LD options (**Supplementary Note Tables 1 and 2**) than a GWS locus versus a 3Mb fine-mapping window (**Supplementary Note Tables 3 and 4**). For the full PGC3 BD GWAS, we ran fine-mapping using the four LD options to ensure the robustness of fine-mapped SNPs. In the full PGC3 BD GWAS, 3 of the GWS loci were larger than 3Mb, so could not be fully covered by a 3Mb window. The remaining 61 GWS loci varied greatly in size, ranging from 14,960 bp - 1.96 Mb (mean 375 kb (sd 390 kb). Based on these results, we fine-mapped the GWS loci from the full PGC3 BD GWAS using only the GWS locus fine-mapping windows (defined in the original GWAS as LD $r^2 > 0.1$ with the top lead SNP within a 3 Mb range), as these best reflect the regions of genome-wide significant association from the GWAS. Additionally, using a fine-mapping window that is significantly larger than the GWS locus window, may fine-map other nearby regions that capture different association signals to that represented by the GWS index SNP.

Description of testing cohorts used for polygenic risk scoring analyses

Below we describe the ascertainment and diagnosis of the participants in each testing cohort. Most cohorts have been published individually, and the primary report can usually be found using the PubMed identifiers provided. The lead PI of each sample warranted that their protocol was approved by their local Ethical Committee and that all participants provided written informed consent. The boldfaced first line for each sample indicates study PI, PubMed ID if published, country (study name), the Psychiatric Genomics Consortium internal tag or study identifier, and genetic ancestry.

Grigoroiu-Serbanescu M | PMID : 31791676| Romania (BOMA-Romania) | rom4 | European

Patient sample. Unrelated bipolar disorder type I (BD-I) patients (N=102) were recruited from consecutive admissions in the Obregia Psychiatric Hospital of Bucharest, Romania. All participants provided written informed consent following a detailed explanation of the study aims and procedures. The study was performed in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) and it was approved by the ethical committee of the hospital and by the reviewers of the Research and Education Ministry. All participants were of Romanian descent according to self-reported ancestry. Genealogical information about parents and all four grandparents was obtained through direct interview of the subjects. The patients were investigated with the Diagnostic Interview for Genetic Studies (DIGS)[2] and the Family Interview for Genetic Studies (FIGS)[3]. Information was also obtained from medical records and close relatives. The diagnosis of BD-I was assigned according to DSM-IV-TR criteria using the best estimate procedure that considered patient interview, medical records and information provided by close relatives. Patients were included in the sample if they had at least two documented hospitalized illness episodes (one manic/mixed and one depressive or two manic episodes) and no residual mood incongruent psychotic symptoms during remissions. This information was also confirmed by first degree relatives for 78% of the cases in face to face interviews. Family history of psychiatric illness was obtained with FIGS administered both to the patients and to all available first degree relatives.

Control sample. The controls (N=198) were volunteers from the personnel and students of the University of Bucharest, as well as from personnel and medical residents at the Obregia Psychiatric Hospital and the Institute of Virology of Bucharest. All controls were evaluated using the DIGS and FIGS to screen for a lifetime history of major affective disorders, schizoaffective disorders, SCZ and other psychoses, obsessive-compulsive disorder, eating disorders, and alcohol or drug addiction. Unaffected individuals were included as controls in the present study.

Genotyping of the Romanian patients and controls. The BD-I cases and controls were genome-wide genotyped on Illumina GSA-MD beadchips. Stringent quality control was applied to the genotype information. Individuals were excluded on the basis of having incorrect gender assignments; excessive heterozygosity (more than 10 standard deviations above the mean); missing genotype data above 10% and evidence of relatedness. SNPs were excluded with a minor allele frequency < 0.5% and deviating substantially from the Hardy-Weinberg equilibrium ($P < 10^{-6}$).

Kircher T | PMID 30267149| Germany | FOR 2107 |European

The FOR2107 cohort is a multi-centre study, recruited through newspaper advertisements and mailing lists from the areas of Marburg and Muenster in Germany[4]. The sample includes 147 cases and 696

controls, genotyped with the PsychChip platform. Ethics approval was obtained from the ethics committees of the Medical Schools of the Universities of Marburg and Muenster, respectively, in accordance with the Declaration of Helsinki. All subjects volunteered to participate in the study and provided written informed consent.

McQuillin A | PMID: 37643680 | UCL (University College London), London, UK | amq1| European

The Amq1 cohort is a study taking place in London, United Kingdom. Diagnostic criteria were based on ICD-10 codes and clinical interviews and controls were screened for any mental disorder. The cohort includes 417 cases (201 bipolar disorder type I (BD-I) and 39 bipolar disorder type II (BD-II) cases) and 533 control samples, genotyped using the A5.0 platform. The sample was composed of Caucasian individuals who were ascertained and received clinical diagnoses of BD-I according to UK National Health Service (NHS) psychiatrists at interview using ICD-10 codes. In addition bipolar subjects were included only if both parents were of English, Irish, Welsh or Scottish descent and if three out of four grandparents were of the same descent. All volunteers read an information sheet approved by the Metropolitan Medical Research Ethics Committee who also approved the project for all NHS hospitals. Written informed consent was obtained from each volunteer. The control subjects were recruited from London branches of the National Blood Service, from local NHS family doctor clinics and from university student volunteers. All control subjects were interviewed with the SADS-L to exclude all psychiatric disorders.

Pato M, Pato C, Bigdeli T | PMID: 33169155 | USA | GPC | Admixed African American, Latino

Details of ascertainment and diagnosis, genotyping and quality control have been described in detail previously[5]. Briefly, cases were ascertained using the Diagnostic Interview for Psychosis and Affective Disorders (DI-PAD), a semi-structured clinical interview administered by mental health professionals, which was developed specifically for the GPC study. Individuals reporting no lifetime symptoms indicative of psychosis or mania and who have no first-degree relatives with these symptoms are included as control participants. Genotyping of the cohort was performed in 7 ‘batches’ using Illumina Infinium arrays (Omni2.5, Multi-Ethnic Global Array, and Global Screening Array). Typed variants were aligned to the human reference genome (GRCh37), and within each genotyping batch, variants with missingness greater than 2% or Hardy-Weinberg Equilibrium *P*-value<10-6 were excluded; all scripts for pre-processing GWAS array data are downloadable from https://github.com/freeseek/gwaspipeline.

Computational phasing and statistical genotype imputation were performed for each genotyping batch using Eagle (v2.3.5) and Minimac3 (v2.0.1), respectively, with default parameters and using publicly available reference haplotypes from the 1000 Genomes Project (1KGP) Phase 3. Principal components analysis (PCA) was performed with GCTA (v1.2.4), using a genome-wide genetic relatedness matrix (GRM) estimated for the full GPC dataset and reference samples from the 1KGP Phase 3 data based on 34,918 genotyped SNPs. For each individual, we estimated genome-wide average proportions of African (AFR), European (EUR), Admixed American (AMR), East Asian (EAS), and South Asian (SAS) ancestry from global ancestry PCs using a simple linear mixed model. The African American GPC cohort included 1766 cases and 2535 controls, while the Latino GPC cohort comprised 1032 cases and 3090 controls.

Iwata N | PMID: 28115744 | Japan (advanced COSMO and Biobank Japan) | East Asian
A detailed description of the sample information, genotyping, quality control and imputation procedures is reported elsewhere[6]. In brief, 2,964 BD and 61,887 comparison subjects from the Japanese population were included in this dataset (genotyped by Illumina OmniExpressExome v1.0 or v1.2 BeadChips). After the imputation and stringent QC, a total of 6,195,093 imputed SNPs were analyzed for the association analysis. The diagnosis for each case subject followed the DSM-IV-TR criteria for BD and schizoaffective disorder and was reached by the consensus of at least two experienced psychiatrists, based on unstructured interviews with the subject and their family, as well as a review of the subject's medical records. For the comparison subjects, we used GWAS data for subjects in the BioBank Japan project collected as case subjects for non-psychiatric disorders. These subjects were not psychiatrically evaluated.

Hong-Hee Won, Woojae Myung, Heon-Jeong Lee, Genoplan Research Team | Not published | South Korea | East Asian
We genotyped 807 patients with bipolar disorder, 726 patients with schizophrenia and 497 healthy control subjects using the Affymetrix AxiomR Korea Biobank Array 1.0 (K-CHIP). K-CHIP was designed by the Center for Genome Science at the Korea National Institute of Health, including 833K SNPs. A more detailed description of the genotyping procedure is reported elsewhere. We performed sample-level and variant-level QC of genotype data. We excluded variants with missing rate > 1%, Hardy-Weinberg equilibrium $P < 10^{-6}$, or minor allele frequency < 1%, and samples with missing rate > 5%, relatedness among the sample, mismatch between self-reported and inferred sex, or deviated heterozygosity rate. We confirmed homogeneity of the samples based on visual inspection of principal component analysis plots. Genotype imputation was conducted using the Haplotype Reference Consortium (HRC) reference panel. After the imputation and additional post-QC ($R^2 > 0.8$ and minor allele frequency > 1%), a total of 770 bipolar cases and 497 controls and 5,483,856 variants were analyzed for polygenic risk score. All the patients met the DSM-IV-TR diagnostic criteria for bipolar I disorder and bipolar II disorder. For clinical diagnosis, a structured interview using the Korean version of the Diagnostic Interview for Genetic Studies (DIGS) or the Structured Clinical Interview for DSM-IV (SCID) was performed. The control group consisted of volunteers from the community who were free of any history of clinically significant psychiatric symptoms. Detailed assessment processes are described elsewhere[7].

Full Acknowledgments

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.

Investigator Acknowledgements

NM and MK were part funded by the Baszucki Brain Research Fund via the Milken Institute Center for Strategic Philanthropy and the NIMH US National Institute of Mental Health (R01MH124839).

JRIC is part funded by the NIHR Maudsley Biomedical Research Centre. This paper represents independent research funded 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 UK NIHR or Department of Health and Social Care. JRIC is part funded by a grant from the UK Medical Research Foundation (MRF-001-0012-RG-COLE-C0930).

Cohort Acknowledgements

BACCS: This work was supported in part by the NIHR Maudsley Biomedical Research Centre ('BRC') hosted at King's College London and South London and Maudsley NHS Foundation Trust, and funded by the National Institute for Health Research under its Biomedical Research Centres funding initiative. The views expressed are those of the authors and not necessarily those of the BRC, the NHS, the NIHR or the Department of Health or King's College London. We gratefully acknowledge capital equipment funding from the Maudsley Charity (Grant Reference 980) and Guy's and St Thomas's Charity (Grant Reference STR130505).

BACCS-Canada: This work was supported through funding from the Canadian Institutes of Health Research, MOP-172013, to JBV at the Centre for Addiction & Mental Health, Toronto. The ascertainment of the case control cohorts was also supported by funding from GlaxoSmithKline to JBV.

BD_TRS: This work was funded by the German Research Foundation (DFG, grant FOR2107 DA1151/5-1 to UD; SFB-TRR58, Project C09 to UD) and the Interdisciplinary Center for Clinical Research (IZKF) of the medical faculty of Münster (grant Dan3/012/17 to UD).

BiGS, GAIN: FJM was supported by the NIMH Intramural Research Program, NIH, DHHS.

BOMA-Australia: This work was supported by the Australian National Health and Medical Research Council, grant numbers: 1037196 (PBM, PRS), 1066177 (JMF, JIN), 1063960 (JMF, PRS); and the Lansdowne Foundation. BOMA-Germany I, BOMA-Germany II, BOMA-Germany III, PsyCourse: This work was supported by the German Ministry for Education and Research (BMBF) through the Integrated Network IntegraMent (Integrated Understanding of Causes and Mechanisms in Mental Disorders), under the auspices of the e:Med program (grant 01ZX1314A/01ZX1614A to MMN and SC, grant 01ZX1314G/01ZX1614G to MR, grant 01ZX1314K to TGS). This work was supported by the

German Ministry for Education and Research (BMBF) grants NGFNplus MooDS (Systematic Investigation of the Molecular Causes of Major Mood Disorders and Schizophrenia; grant 01GS08144 to MMN and SC, grant 01GS08147 to MR). This work was also supported by the Deutsche Forschungsgemeinschaft (DFG), grant NO246/10-1 to MMN (FOR 2107), grant RI 908/11-1 to MR (FOR 2107), grant WI 3429/3-1 to SHW, grants SCHU 1603/4-1, SCHU 1603/5-1 (KFO 241) and SCHU 1603/7-1 (PsyCourse) to TGS. This work was supported by the Swiss National Science Foundation (SNSF, grant 156791 to SC). MMN is supported through the Excellence Cluster ImmunoSensation. TGS is supported by an unrestricted grant from the Dr. Lisa-Oehler Foundation. AJF received support from the BONFOR Programme of the University of Bonn, Germany. MH was supported by the Deutsche Forschungsgemeinschaft.

Fran (France): This research was supported by Assistance Publique des Hôpitaux de Paris (APHP Grant PHRC GAN12), by Institut National de la Santé et de la Recherche Médicale (INSERM grant C0829), by the Fondation FondaMental and by the Investissements d'Avenir Programs managed by the Agence nationale pour la Recherche (references ANR-11-IDEX-0004-02 and ANR-10-COHO-10-01).

Genomic Psychiatry Cohort: The GPC was supported by grants R01 MH085548, R01 MH104964, R01 R01MH104964, and R01MH123451 from the National Institute of Mental Health (NIMH), and genotyping of samples was provided by the Stanley Center for Psychiatric Research at Broad Institute. Funding support for the Whole Genome Association Study of Bipolar Disorder and the Genome-Wide Association of Schizophrenia Study was provided by the NIMH (R01 MH67257, R01 MH59588, R01 MH59571, R01 MH59565, R01 MH59587, R01 MH60870, R01 MH59566, R01 MH59586, R01 MH61675, R01 MH60879, R01 MH81800, U01 MH46276, U01 MH46289, U01 MH46318, U01 MH79469, and U01 MH79470) and the genotyping of samples was provided through the Genetic Association Information Network (GAIN). The following investigators contributed to GPC cohorts:
Michele T Pato MD [44], Carlos N Pato, MD, PhD [44], Tim B Bigdeli, PhD [1,2,3], Ayman H Fanous, MD [45,46,47], Steven A McCarroll, PhD [4,5], Peter F Buckley, MD [6], Mark J. Daly [7,8,9,5], James A Knowles MD, PhD [2,10], Douglas S Lehrer, MD [11], Dolores Malaspina, MD, MSPH [12,13], Mark H Rapaport, MD [14], Jeffrey J Rakofsky, MD [14], Janet L Sobell, PhD [15], Giulio Genovese, PhD [4,5], Penelope Georgakopoulos, DrPH [2], Jacquelyn L Meyers, PhD [1], Roseann E Peterson, PhD [6], Helena Medeiros, MSW [2], Jorge Valderrama, PhD [1,2], Eric D Achtyes, MD [16], Roman Kotov, PhD [17], Colony Abbott, MPH [16], Maria Helena Azevedo, PhD [18], Richard A Belliveau, Jr, BA [4], Elizabeth Bevilacqua, BS [19], Evelyn J Bromet, PhD [17], William Byerley, MD [20], Celia Barreto Carvalho, PhD [21], Sinéad B Chapman, MS [4], Lynn E DeLisi, MD [22,23], Ashley L Dumont, BASc [4], Colm O'Dushlaine, PhD [4], Laura J Fochtmann, MD [17], Diane Gage [4], James L Kennedy, MD [24], Becky Kinkead, PhD [14], Antonio Macedo, PhD [18], Jennifer L Moran, PhD [4], Christopher P Morley, PhD [25-27], Mantosh J Dewan, MD [27], James Nemesh [4], Diana O Perkins, MD, MPH [28], Shaun M Purcell, PhD [4,29], Edward M Scolnick, MD [4], Brooke M Sklar, MA [15], Pamela Sklar, MD, PhD [12,13], Jordan W Smoller, MD, ScD [4,23,30,31], Patrick F Sullivan, MD, FRANZCP [28,32], Humberto Nicolini, MD [33], Conrad O Iyegbe, PhD [34], Fabio Macciardi, MD, PhD [35], Stephen R Marder, MD [36,37], Michael A Escamilla, MD [38], Ruben C Gur, PhD [39-41], Raquel E Gur, MD, PhD [39-41], Tiffany A Greenwood, PhD [42], David L Braff, MD [42,43], Marquis P Vawter, PhD, MA, MS [35], Chris Chatzinakos, PhD [1,2]

[1] Department of Psychiatry and Behavioral Sciences and [2] Institute for Genomics in Health, SUNY Downstate Medical Center, Brooklyn, NY, USA; [3] VA New York Harbor Healthcare System, Brooklyn, NY, USA; [4] Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA; [5] Department of Genetics, Harvard Medical School, Boston, MA, USA; [6] School of Medicine, Virginia Commonwealth University, Richmond, VA, USA; [7] Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; [8] Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; [9] Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA; [10] Department of Cell Biology, SUNY Downstate Medical Center, Brooklyn, NY, USA; [11] Department of Psychiatry, Wright State University, Dayton, OH, USA; [12] Departments of Psychiatry and [13] Genetics & Genomics, Icahn School of Medicine at Mount Sinai, NY, USA; [14] Department of Psychiatry and Behavioral Sciences, Emory University, Atlanta, GA, USA; [15] Department of Psychiatry & Behavioral Sciences, University of Southern California, Los Angeles, CA, USA; [16] Cherry Health and Michigan State University College of Human Medicine, Grand Rapids, MI, USA; [17] Department of Psychiatry, Stony Brook University, Stony Brook, NY, USA; [18] Institute of Medical Psychology, Faculty of Medicine, University of Coimbra, Coimbra, PT; [19] Beacon Health Options, Boston, MA, USA; [20] Department of Psychiatry, University of California, San Francisco, CA, USA; [21] Faculty of Social and Human Sciences, University of Azores, PT; [22] VA Boston Healthcare System, Brockton, MA, USA; [23] Department of Psychiatry, Harvard Medical School, Boston, MA, USA; [24] Neurogenetics Laboratory, Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health; Department of Psychiatry, University of Toronto, ON, CA; [25] Departments of Public Health and Preventive Medicine, [26] Family Medicine, and [27] Psychiatry and Behavioral Sciences, State University of New York, Upstate Medical University, Syracuse, NY, USA; [28] Department of Psychiatry, University of North Carolina, Chapel Hill, NC, USA; [29] Department of Psychiatry, Brigham and Women's Hospital, Boston, MA, USA; [30] Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA; [31] Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA; [32] Medical Epidemiology and Biostatistics, Karolinska Institutet, Solna, SE; [33] Carracci Medical Group, Mexico City, MX; [34] Department of Psychosis Studies, King's College London, London, UK; [35] Department of Psychiatry and Human Behavior, University of California, Irvine, CA, USA; [36] Department of Psychiatry and Biobehavioral Sciences and [37] Semel Institute for Neuroscience and Human Behavior, Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA; [38] Department of Psychiatry, University of Texas Rio Grande Valley School of Medicine; [39] Departments of Psychiatry and [40] Child & Adolescent Psychiatry and [41] Lifespan Brain Institute, University of Pennsylvania Perelman School of Medicine and Children's Hospital of Philadelphia, Philadelphia, PA, USA; [42] Department of Psychiatry, University of California, La Jolla, San Diego, CA, USA; [43] VISN-22 Mental Illness, Research, Education and Clinical Center (MIRECC), VA San Diego Healthcare System, San Diego, CA, USA; [44] Robert Wood Johnson Medical School, Psychiatry, Rutgers, NJ, USA; [45] The University of Arizona College of Medicine-Phoenix; [46] Banner-University Medical Center; [47] Carl T. Hayden Veterans Administration Medical Center (Phoenix).

Halifax: Halifax data were obtained with support from the Canadian Institutes of Health Research (grant #166098), Genome Canada and from Dalhousie Medical Research Foundation.

The Mayo Bipolar Disorder Biobank was funded by the Marriot Foundation and the Mayo Clinic Center for Individualized Medicine.

Michigan (NIMH/Pritzker Neuropsychiatric Disorders Research Consortium): We thank the participants who donated their time and DNA to make this study possible. We thank members of the NIMH Human Genetics Initiative and the University of Michigan Prechter Bipolar DNA Repository for generously providing phenotype data and DNA samples. Many of the authors are members of the Pritzker Neuropsychiatric Disorders Research Consortium which is supported by the Pritzker Neuropsychiatric Disorders Research Fund L.L.C. A shared intellectual property agreement exists between this philanthropic fund and the University of Michigan, Stanford University, the Weill Medical College of Cornell University, HudsonAlpha Institute of Biotechnology, the Universities of California at Davis, and at Irvine, to encourage the development of appropriate findings for research and clinical applications.

Neuc1 (NeuRA-BDHR-Australia): This work was supported by the Australian National Health and Medical Research Council, grant numbers: 1037196 (PBM, PRS), 1066177 (JMF, JIN), 1063960 (JMF, PRS); and the Lansdowne Foundation. JMF would like to thank Janette M O'Neil and Betty C Lynch for their support.

Neuc1 (NeuRA-CASSI-Australia): This work was funded by the NSW Ministry of Health, Office of Health and Medical Research. CSW was a recipient of National Health and Medical Research Council (Australia) Fellowships (#1117079, #1021970).

Neuc1 (NeuRA-IGP-Australia): MJG was supported by a NHMRC Career Development Fellowship (1061875).

Neuc1/ASGC1/ASRB: This study used samples and data from the Australian Schizophrenia Research Bank (ASRB), which is supported by the National Health and Medical Research Council of Australia, the Pratt Foundation, Ramsay Health Care, the Viertel Charitable Foundation and the Schizophrenia Research Institute. We thank and acknowledge the contribution of the ASRB Chief Investigators: V. Carr, U. Schall, R. Scott, A. Jablensky, B. Mowry, P. Michie, S. Catts, F. Henskens and C. Pantelis.

Span2: CSM was a recipient of a Sara Borrell contract (CD15/00199) and a mobility grant (MV16/00039) from the Instituto de Salud Carlos III, Ministerio de Economía, Industria y Competitividad, Spain. MR was a recipient of a Miguel de Servet contract (CP09/00119 and CPII15/00023) from the Instituto de Salud Carlos III, Ministerio de Economía, Industria y Competitividad, Spain. This investigation was supported by Instituto de Salud Carlos III (PI14/01700, PI15/01789, PI16/01505,PI17/00289, PI18/01788, PI19/00721, and P19/01224), and cofinanced by the European Regional Development Fund (ERDF), “la Marató de TV3” (092330/31), the Agència de Gestió d’Ajuts Universitaris i de Recerca-AGAUR, Generalitat de Catalunya (2014SGR1357 and 2017SGR1461) and the Pla estratègic de recerca i innovació en salut (PERIS), Generalitat de Catalunya (MENTAL-Cat; SLT006/17/287). This project has also received funding from the European Union’s Horizon 2020 Research and Innovation Programme under the grant agreements No 667302 (CoCA) and 728018 (Eat2beNICE).

SWEBIC: We are deeply grateful for the participation of all subjects contributing to this research, and to the collection team that worked to recruit them. We also wish to thank the Swedish National Quality Register for Bipolar Disorders: BipoläR. Funding support was provided by the Stanley Center for Psychiatric Research, Broad Institute from a grant from Stanley Medical Research Institute, the Swedish Research Council, and the NIMH.

Sweden: This work was funded by the Swedish Research Council (M. Schalling, C. Lavebratt), the Stockholm County Council (M. Schalling, C. Lavebratt, L. Backlund, L. Frisén, U. Ösby) and the Söderström Foundation (L. Backlund) and the Swedish Brain Foundation (T. Olsson).

UK - BDRN: BDRN would like to acknowledge funding from the Wellcome Trust and Stanley Medical Research Institute, and especially the research participants who continue to give their time to participate in our research.

UNIBO / University of Barcelona, Hospital Clinic, IDIBAPS, CIBERSAM: EV thanks the support of the Spanish Ministry of Economy and Competitiveness (PI15/00283) integrated into the Plan Nacional de I+D+I y cofinanciado por el ISCIII-Subdirección General de Evaluación y el Fondo Europeo de Desarrollo Regional (FEDER); CIBERSAM; and the Comissionat per a Universitats i Recerca del DIUE de la Generalitat de Catalunya to the Bipolar Disorders Group (2014 SGR 398).

WTCCC: The principal funder of this project was the Wellcome Trust. For the 1958 Birth Cohort, venous blood collection was funded by the UK Medical Research Council.

Funding sources

Study	**Lead investigator**	**Country, Funder, Award number**
Statistical and functional fine-mapping of bipolar disorder genetic risk loci	N Mullins	USA, Baszucki Brain Research Fund
PGC	PF Sullivan; EA Stahl	USA, NIMH MH109528; NIMH U01 MH109536
PGC	D Posthuma	Netherlands, Scientific Organization Netherlands, 480-05-003
PGC	D Posthuma	Dutch Brain Foundation and the VU University Amsterdam Netherlands

UK - BDRN	N Craddock I Jones L Jones MJ Owen	Medical Research Council (MRC) Centre (MR/P005748/1G0801418) and Program Grants (MR/P005748/1G0800509)
ASRB	V Carr	Australia, National Health and Medical Research Council, grant number: (86500).
ASRB	M Cairns	Australia, National Health and Medical Research Council, grant numbers: (1121474, 1147644).
ASRB	C Pantelis	Australia, National Health and Medical Research Council of Australia (grants IDs: 1196508, 1150083).
BACCS	G Breen	GB, JRIC, HG, CL were supported in part by the NIHR Maudsley Biomedical Research Centre ('BRC') hosted at King's College London and South London and Maudsley NHS Foundation Trust, and funded by the National Institute for Health Research under its Biomedical Research Centres funding initiative.
BD_TRS	U Dannlowski	Germany, DFG, Grant FOR2107 DA1151/5-1; Grant SFB-TRR58, Project C09
BiGS, Uchicago	ES Gershon	R01 MH103368
BiGS, NIMH	FJ McMahon	US, NIMH, R01 MH061613, ZIA MH002843
BiGS, GAIN, UCSD	J Kelsoe	US, NIMH, MH078151, MH081804, MH59567
BOMA-Australia	JM Fullerton	Australia, National Health and Medical Research Council, grant numbers: 1066177; 1063960
BOMA-Australia	SE Medland	Australia, National Health and Medical Research Council, grant numbers: 1103623
BOMA-Australia	PB Mitchell	Australia, National Health and Medical Research Council, grant numbers: 1037196

BOMA-Australia	GW Montgomery	Australia, National Health and Medical Research Council, grant numbers: 1078399
BOMA-Australia	PR Schofield	Australia, National Health and Medical Research Council, grant numbers: 1037196; 1063960; 1176716
ROMANIA rom4_eur (BOMA-Romania; Bip_rom3_eur, bmrom)	M Grigoroiu-Serbanescu	ROMANIA, UEFISCDI, Romania, Grant nr. 203/2021 (code PN-III-P4-ID-PCE-2020-2269)
BOMA-Germany I, II, III	S Cichon	Germany, BMBF Integrament, 01ZX1314A/01ZX1614A
BOMA-Germany I, II, III	S Cichon	Germany, BMBF NGFNplus MooDS, 01GS08144
BOMA-Germany I, II, III	S Cichon	Switzerland, SNSF, 156791
BOMA-Germany I, II, III	S Cichon	Switzerland, SNSF, 182731
BOMA-Germany I, II, III	MM Nöthen	Germany, BMBF Integrament, 01ZX1314A/01ZX1614A
BOMA-Germany I, II, III	MM Nöthen	Germany, BMBF NGFNplus MooDS, 01GS08144
BOMA-Germany I, II, III	MM Nöthen	Germany, Deutsche Forschungsgemeinschaft, Excellence Cluster ImmunoSensation
BOMA-Germany I, II, III	MM Nöthen	Germany, Deutsche Forschungsgemeinschaft, NO246/10-1
BOMA-Germany I, II, III	SH Witt	Germany, Deutsche Forschungsgemeinschaft, WI 3429/3-2
BOMA-Germany I, II, III, BOMA-Spain	M Rietschel	Germany, Deutsche Forschungsgemeinschaft, RI 908/11-1

BOMA-Germany I, II, III, BOMA-Spain	M Rietschel	Germany, BMBF, ERA-Net Neuron “EMBED”, 01EW1904
BOMA-Germany I, II, III, BOMA-Spain	M Rietschel	Germany, BMBF, ERA-Net Neuron “Synschiz”, 01EW1810
BOMA-Germany I, II, III, PsyCourse, BiGS	TG Schulze	Germany, BMBF Integrament, 01ZX1314K
BOMA-Germany I, II, III, PsyCourse, BiGS	TG Schulze	Germany, DFG, SCHU 1603/4-1, SCHU 1603/5-1, SCHU 1603/7-1
BOMA-Germany I, II, III, PsyCourse, BiGS	TG Schulze	Germany, Dr. Lisa-Oehler Foundation (Kassel, Germany)
Bulgarian Trios (Cardiff)	G Kirov MJ Owen	The recruitment was funded by the Janssen Research Foundation. Genotyping was funded by multiple grants to the Stanley Center for Psychiatric Research at the Broad Institute from the Stanley Medical Research Institute, The Merck Genome Research Foundation, and the Herman Foundation.
FOR2107	T Kircher, U Dannlowski, I Nenadić	German Research Foundation (Deutsche Forschungsgemeinschaft, DFG grant nos. KI 588/14-1, KI 588/14-2, KR 3822/7-1, KR 3822/7-2, NE 2254/1-2, NE 2254/2-1, NE2254/3-1, NE2254/4-1, DA 1151/5-1, DA 1151/5-2, SCHW 559/14-1, 545/7-2, RI 908/11-2, WI 3439/3-2, NO 246/10-2, DE 1614/3-2, HA 7070/2-2, JA 1890/7-1, JA 1890/7-2, MU 1315/8-2, RE 737/20-2, KI 588/17-1)
France	M Leboyer, F Bellivier, B Etain, S Jamain	France, APHP, INSERM, ANR, Fondation Fondamental
Halifax	M Alda	CIHR grant #166098, Research Nova Scotia, Genome Canada, and Dalhousie Medical Research Foundation
IMAGE	B Franke	National Institutes of Health (grant R01MH62873), Dutch NWO Large Investment Program (grant 1750102007010)

Mayo Bipolar Disorder Biobank	JM Biernacka, MA Frye	Marriot Foundation and the Mayo Clinic Center for Individualized Medicine
Michigan	M Boehnke, RM Myers	US, NIMH, R01 MH09414501A1
Michigan	M Boehnke	US, NIMH, MH105653
Michigan	L Scott	US, NIMH U01 MH085513-01 (sub-contract PI)
Mount Sinai, STEP-BD, FAST	P Sklar, EA Stahl	US NIH R01MH106531, R01MH109536
NeuRA-CASSI-Australia (neuc1)	C Shannon Weickert	Australia, National Health and Medical Research Council, grant number: 568807
NeuRA-CASSI-Australia (neuc1)	TW Weickert	Australia, National Health and Medical Research Council, grant number: 568807
NeuRA-IGP-Australia (neuc1)	MJ Green	Australia, National Health and Medical Research Council, grant numbers: 630471, 1081603
NeuRA-BDHR-Australia (neuc1)	PB Mitchell, PR Schofield, JM Fullerton	Australia, National Health and Medical Research Council, grant numbers: 1037196; 1066177; 1063960, 1200428, 1176716, 1177991 & The Lansdowne Foundation
Norway	OA Andreassen	Norway, Research Council of Norway (#223273, #248778, #249711, #273291, #296030,#324499 #324252), KG Jebsen Stiftelsen, The South-East Norway Regional Health Authority (#2017-004, #2022-073, #2023-031) EU's H2020 RIA grant # 964874 REALMENT
Norway	KS O'Connell	US NIH 5R01MH124839-02, Research Council of Norway (#334920)
Span2	M Ribasés	Instituto de Salud Carlos III (PI19/01224, PI20/00041, PI22/00464), "la Marató de TV3" (202228-30 and 202228-31), the Agència de Gestió d'Ajuts Universitaris i de Recerca-AGAUR, Generalitat de Catalunya (2021SGR-00840), Fundació 'la Caixa, Diputació de Barcelona, Pla Estratègic de

		Recerca i Innovació en Salut (PERISSLT006/17/285) and Fundació Privada d'Investigació Sant Pau(FISP)
State University of New York, Downstate Medical Center (SUNY DMC)	C Pato, MT Pato, JA Knowles, H Medeiros	US, National Institutes of Health, R01MH085542
SWEBIC	M Landén	The Stanley Center for Psychiatric Research, Broad Institute from a grant from Stanley Medical Research Institute, the Swedish Research Council (2022-01643), the Swedish foundation for Strategic Research (KF10-0039), and the Swedish Brain foundation (FO2022-0217).
UCL	A McQuillin	Medical Research Council (MRC) - G1000708
UCLA-Utrecht (Los Angeles)	RA Ophoff	US, National Institutes of Health, R01MH090553, R01MH115676
UK - BDRN (Cardiff)	MC O'Donovan	Medical Research Council (MRC) Centre (MR/L010305/1G0801418) and Program Grants (MR/P005748/1G0800509)
UK - BDRN (Cardiff)	MJ Owen	Medical Research Council (MRC) Centre (MR/L010305/1G0801418) and Program Grants (MR/P005748/1G0800509)
UK - BDRN (Cardiff and Worcester)	N Craddock, I Jones, LA Jones	UK, Wellcome Trust, 078901; USA, Stanley Medical Research Institute, 5710002223-01
UK - BDRN (Cardiff)	A Di Florio	European Commission Marie Curie Fellowship, grant number 623932.
UNIBO / University of Barcelona, Hospital Clinic, IDIBAPS, CIBERSAM	E Vieta	Grants PI15/00283 (Spain) and 2021 SGR 01358 (Catalonia)

University of Pittsburgh	V Nimgaonkar	US, NIMH MH63480
USC	JL Sobell	USA, National Institutes of Health, R01MH085542
WTCCC	N Craddock; AH Young	Wellcome Trust. For the 1958 Birth Cohort, venous blood collection was funded by the UK Medical Research Council. AHY was funded by NIMH (USA); CIHR (Canada); NARSAD (USA); Stanley Medical Research Institute (USA); MRC (UK); Wellcome Trust (UK); Royal College of Physicians (Edin); BMA (UK); UBC-VGH Foundation (Canada); WEDC (Canada); CCS Depression Research Fund (Canada); MSFHR (Canada); NIHR (UK); Janssen (UK)
Greece	G.P. Patrinos	European Commission (H2020-668353; Ubiquitous Pharmacogenomics (U-PGx); European Commission (Horizon Europe-101057639; SafePolyMed); Greek General Secretariat of Research and Technology (MIS 5002550)
University of California, Irvine	MP Vawter	USA, National Institutes of Health, MH113177, MH074307, MH085801, MH099440, RR000827, MH060068, MH060870. Pritzker Neuropsychiatric Disorders Research Fund L.L.C.
Korea	HJ Lee, Woojae Myung, Hong-Hee Won	National Research Foundation of Korea (2019R1A2C2084158 and 2017M3A9F1031220 to HJ Lee; NRF-2021R1A2C4001779 to WM;NRF-2022R1A2C2009998 to HHW). Ministry of Health & Welfare of Korea (HM14C2606)
BIPGRAZ, Austria	EZ Reininghaus	Funding from the government of Styria, Austria.
BIPGRAZ, Austria	SA Bengesser	MEFO funding of the Medical University of Graz.
Japan	N Iwata, M Ikeda	Jakpan Agency for Medical Research and Development (JP20dm0107097, JP20km0405201, JP20km0405208)
FAST-STEP	JW Smoller	USA, NIH, R01MH063445
ICCBD	JW Smoller, P Sklar	USA, NIH, R01MH085542

Edinburgh, Scotland	A McIntosh	Wellcome funding, reference 220857/Z/20/Z

References

1. 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).
2. Nurnberger, J. I., Jr *et al.* Diagnostic interview for genetic studies. Rationale, unique features, and training. NIMH Genetics Initiative. *Arch. Gen. Psychiatry* **51**, 849–59; discussion 863–4 (1994).
3. Elizabeth Maxwell, M. Family Interview for Genetic Studies (FIGS): a manual for FIGS. *Clinical Neurogenetics Branch, Intramural Research Program, National Institute of Mental Health, Bethesda, MD* (1992).
4. Kircher, T. *et al.* Neurobiology of the major psychoses: a translational perspective on brain structure and function-the FOR2107 consortium. *Eur. Arch. Psychiatry Clin. Neurosci.* **269**, 949–962 (2019).
5. 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).
6. 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).
7. Baek, J. H. *et al.* Psychopathologic structure of bipolar disorders: exploring dimensional phenotypes, their relationships, and their associations with bipolar I and II disorders. *Psychol. Med.* **49**, 2177–2185 (2019).