# Genome-wide association study of 40,000 individuals identifies two novel loci associated with bipolar disorder

Liping Hou[1], Sarah E. Bergen[2,3], Nirmala Akula[1], Jie Song[2], Christina M. Hultman[2], Mikael Landén[4,2], Mazda Adli[5], Martin Alda[6], Raffaella Ardau[7], Bárbara Arias[8], Jean-Michel Aubry[9], Lena Backlund[10], Judith A. Badner[11], Thomas B. Barrett[12], Michael Bauer[13], Bernhard T. Baune[14], Frank Bellivier[15], Antonio Benabarre[16], Susanne Bengesser[17], Wade H. Berrettini[18], Abesh Kumar Bhattacharjee[19], Joanna M. Biernacka[20,21], Armin Birner[17], Cinnamon S. Bloss[22], Clara Brichant-Petitjean[15], Elise T. Bui[1], William Byerley[23], Pablo Cervantes[24], Caterina Chillotti[7], Sven Cichon[25,26,27,28], Francesc Colom[16], William Coryell[29], David W. Craig[30], Cristiana Cruceanu[31], Piotr M. Czerski[32], Tony Davis[14], Alexandre Dayer[9], Franziska Degenhardt[25,26], Maria Del Zompo[33], J. Raymond DePaulo[34], Howard J. Edenberg[35], Bruno Étain[36], Peter Falkai[37], Tatiana Foroud[38], Andreas J. Forstner[25,26], Louise Frisén[39,40], Mark A. Frye[21], Janice M. Fullerton[41,42], Sébastien Gard[43], Julie S. Garnham[6], Elliot S. Gershon[11], Fernando S. Goes[34], Tiffany A. Greenwood[19], Maria Grigoroiu-Serbanescu[44], Joanna Hauser[32], Urs Heilbronner[45,46], Stefanie Heilmann-Heimbach[25,26], Stefan Herms[25,26,28], Maria Hipolito[47], Shashi Hitturlingappa[14], Per Hoffmann[25,26,27,28], Andrea Hofmann[25,26], Stephane Jamain[36], Esther Jiménez[16], Jean-Pierre Kahn[48], Layla Kassem[1], John R. Kelsoe[19], Sarah Kittel-Schneider[49], Sebastian Kliwicki[50], Daniel L. Koller[38], Barbara König[51], Nina Lackner[17], Gonzalo Laje[1], Maren Lang[52], Catharina Lavebratt[10], William B. Lawson[47], Marion Leboyer[36], Susan G. Leckband[53], Chunyu Liu[54], Anna Maaser[25,26], Pamela B. Mahon[34], Wolfgang Maier[55], Mario Maj[56], Mirko Manchia[33,57], Lina Martinsson[39], Michael J. McCarthy[58], Susan L. McElroy[59], Melvin G. McInnis[60], Rebecca McKinney[19], Philip B. Mitchell[61], Marina Mitjans[8], Francis M. Mondimore[34], Palmiero Monteleone[56,62], Thomas W. Mühleisen[27,25,26], Caroline M. Nievergelt[19], Markus M. Nöthen[25,26], Tomas Novák[63], John I. Nurnberger, Jr.[64], Evaristus A. Nwulia[47], Urban Ösby[65], Andrea Pfennig[13], James B. Potash[66], Peter Propping[25], Andreas Reif[49], Eva Reininghaus[17], John Rice[67], Marcella Rietschel[52], Guy A. Rouleau[68], Janusz K. Rybakowski[50], Martin Schalling[10], William A. Scheftner[69], Peter R. Schofield[42,70], Nicholas J. Schork[22], Thomas G. Schulze[45,1,34,46,52], Johannes Schumacher[25,26], Barbara W. Schweizer[34], Giovanni Severino[33], Tatyana Shekhtman[19], Paul D. Shilling[19], Christian Simhandl[71], Claire M. Slaney[6], Erin N. Smith[22], Alessio Squassina[33], Thomas Stamm[5], Pavla Stopkova[63], Fabian Streit[52], Jana Strohmaier[52], Szabolcs Szelinger[30], Sarah K. Tighe[66], Alfonso Tortorella[56], Gustavo Turecki[31], Eduard Vieta[16], Julia Volkert[49], Stephanie H. Witt[52], Adam Wright[61], Peter P. Zandi[72], Peng Zhang[60], Sebastian Zollner[60], Francis J. McMahon[1]*

[1] Intramural Research Program, National Institute of Mental Health, National Institutes of Health, U.S. Department of Health & Human Services, Bethesda, MD, United States
[2] Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
[3] Stanley Center for Psychiatric Research, The Broad Institute of MIT and Harvard, Cambridge, MA, United States
[4] Institute of Neuroscience and Physiology, The Sahlgrenska Academy at Gothenburg University, Gothenburg, Sweden
[5] Department of Psychiatry and Psychotherapy, Charité - Universitätsmedizin Berlin, Campus Charité Mitte, Berlin, Germany
[6] Department of Psychiatry, Dalhousie University, Halifax, Canada
[7] Unit of Clinical Pharmacology, Hospital University Agency of Cagliari, Cagliari, Italy
[8] Department of Biologia Animal, Unitat d'Antropologia (Dp. Biología Animal), Facultat de Biologia and Institut de Biomedicina (IBUB), Universitat de Barcelona, CIBERSAM, Barcelona, Spain

[9] Department of Mental Health and Psychiatry, Mood Disorders Unit, Geneva University Hospitals, Geneva, Switzerland
[10] Department of Molecular Medicine and Surgery, Karolinska Institutet and Center for Molecular Medicine, Karolinska University Hospital, Stockholm, Sweden
[11] Department of Psychiatry and Behavioral Neuroscience, University of Chicago, Chicago, IL, United States
[12] Portland Veterans Affairs Medical Center, Portland, OR, United States
[13] Department of Psychiatry and Psychotherapy, University Hospital Carl Gustav Carus, Technische Universität, Dresden, Germany
[14] Discipline of Psychiatry, University of Adelaide, Adelaide, Australia
[15] INSERM UMR-S 1144 - Université Paris Diderot. Pôle de Psychiatrie, AP-HP, Groupe Hospitalier Lariboisière-F. Widal, Paris, France
[16] Bipolar Disorder Program, Institute of Neuroscience, Hospital Clinic, University of Barcelona, IDIBAPS, CIBERSAM, Barcelona, Catalonia, Spain
[17] Special outpatient center for bipolar affective disorder, Medical University of Graz, Graz, Austria
[18] Department of Psychiatry, University of Pennsylvania, Philadelphia, PA, United States
[19] Department of Psychiatry, University of California San Diego, San Diego, CA, United States
[20] Department of Health Sciences Research, Mayo Clinic, Rochester, MN, United States
[21] Department of Psychiatry and Psychology, Mayo Clinic, Rochester, MN, United States
[22] Scripps Translational Science Institute, La Jolla, CA, United States
[23] Department of Psychiatry, University of California at San Francisco, San Francisco, CA, United States
[24] McGill University Health Centre, Mood Disorders Program, Montreal, QC, Canada
[25] Institute of Human Genetics, University of Bonn, Germany
[26] Department of Genomics, Life & Brain Center, University of Bonn, Bonn, Germany
[27] Institute of Neuroscience and Medicine (INM-1), Research Centre Jülich, Germany
[28] Division of Medical Genetics and Department of Biomedicine, University of Basel, Switzerland
[29] University of Iowa Hospitals and Clinics, Iowa City, IA, United States
[30] The Translational Genomics Research Institute, Phoenix, AZ, United States
[31] Douglas Mental Health University Institute, McGill University, Montreal, Canada
[32] Psychiatric Genetic Unit, Poznan University of Medical Sciences, Poznan, Poland
[33] Department of Biomedical Sciences, University of Cagliari, Cagliari, Italy
[34] Department of Psychiatry and Behavioral Sciences, Johns Hopkins School of Medicine, Baltimore, MD, United States
[35] Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN, United States
[36] INSERM U955, Psychiatrie translationnelle, Université Paris Est Créteil, Pôle de Psychiatrie et d'Addictologie, Hôpitaux Universitaires Henri Mondor, Créteil, France
[37] Department of Psychiatry and Psychotherapy, Ludwig-Maximilians-University

Munich, Munich, Germany
[38] Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, United States
[39] Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden
[40] Child and Adolescent Psychiatry Research Center, Stockholm, Sweden
[41] Psychiatric Genetics, Neuroscience Research Australia, Sydney, Australia
[42] School of Medical Sciences, University of New South Wales, Sydney, NSW, 2052, Australia
[43] Service de psychiatrie, Hôpital Charles Perrens, Bordeaux, France
[44] Biometric Psychiatric Genetics Research Unit, Alexandru Obregia Clinical Psychiatric Hospital, Bucharest, Romania
[45] Institute of Psychiatric Phenomics and Genomics, Ludwig-Maximilians-University Munich, Munich, Germany
[46] Department of Psychiatry and Psychotherapy, University Medical Center (UMG), Georg-August University Göttingen, Göttingen, Germany
[47] Department of Psychiatry and Behavioral Sciences, Howard University Hospital, Washington, DC, United States
[48] Service de Psychiatrie et Psychologie Clinique, Centre Psychothérapique de Nancy - Université de Lorraine, Nancy, France
[49] Department of Psychiatry, Psychosomatic Medicine and Psychotherapy, University Hospital Frankfurt, Frankfurt, Germany
[50] Department of Adult Psychiatry, Poznan University of Medical Sciences, Poznan, Poland
[51] Department of Psychiatry and Psychotherapeuthic Medicine, Landesklinikum Neunkirchen, Neunkirchen, Austria
[52] Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, Mannheim, Germany
[53] Department of Pharmacy, VA San Diego Healthcare System, San Diego, CA, United States
[54] Department of Psychiatry, University of Illinois at Chicago, Chicago, IL, United States
[55] Department of Psychiatry, University of Bonn, Germany
[56] Department of Psychiatry, University of Naples SUN, Naples, Italy
[57] Department of Pharmacology, Dalhousie University, Halifax, NS, Canada
[58] Department of Psychiatry, VA San Diego Healthcare System, San Diego, CA, United States
[59] Lindner Center of HOPE, University of Cincinnati College of Medicine, Mason, OH, United States
[60] Department of Psychiatry, University of Michigan, Ann Arbor, MI, United States
[61] School of Psychiatry, University of New South Wales, and Black Dog Institute, Sydney, Australia
[62] Neurosciences Section, Department of Medicine and Surgery, University of Salerno, Salerno, Italy
[63] National Institute of Mental Health, Klecany, Czech Republic
[64] Department of Psychiatry, Indiana University School of Medicine, Indianapolis, IN, United States

[65] Department of Neurobiology, Care sciences, and Society, Karolinska Institutet and Center for Molecular Medicine, Karolinska University Hospital, Stockholm, Sweden
[66] Department of Psychiatry, Carver College of Medicine, University of Iowa School of Medicine, Iowa City, IA, United States
[67] Department of Psychiatry, Washington University School of Medicine in St. Louis
[68] Montreal Neurological Institute and Hospital, McGill University, Montreal, Canada
[69] Rush University Medical Center, Chicago, IL, United States
[70] Mental Illness, Neuroscience Research Australia, Sydney, Australia
[71] Bipolar Center Wiener Neustadt, Wiener Neustadt, Austria
[72] Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

***Corresponding author:**

Francis J. McMahon, MD

Chief, Human Genetics Branch
National Institute of Mental Health Intramural Research Program
Email: mcmahonf@mail.nih.gov
Tel: 301-451-4455 / Fax: 301-402-7094
35 Convent Drive, Rm 1A201
Bethesda, MD 20892-3719

## ABSTRACT

Bipolar disorder (BD) is a genetically complex mental illness characterized by severe oscillations of mood and behavior. Genome-wide association studies (GWAS) have identified several risk loci that together account for a small portion of the heritability. To identify additional risk loci, we performed a two-stage meta-analysis of >9 million genetic variants in 9,784 bipolar disorder patients and 30,471 controls, the largest GWAS of BD to date. In this study, to increase power we used ~2,000 lithium-treated cases with a long-term diagnosis of BD from the Consortium on Lithium Genetics, excess controls, and analytic methods optimized for markers on the X-chromosome. In addition to four known loci, results revealed genome-wide significant associations at two novel loci: an intergenic region on 9p21.3 (rs12553324, $p = 5.87\times10^{-9}$; odds ratio = 1.12) and markers within *ERBB2* (rs2517959, $p = 4.53\times10^{-9}$; odds ratio = 1.13). No significant X-chromosome associations were detected and X-linked markers explained very little BD heritability. The results add to a growing list of common autosomal variants involved in BD and illustrate the power of comparing well-characterized cases to an excess of controls in GWAS.

## INTRODUCTION

Bipolar disorder (BD) is a common, chronic and an episodic mental disorder characterized by disruptive oscillations of mood and behaviour. The lifetime prevalence estimate of the US population is about 2% (BD-I and BD-II), but exceeds 2% for sub-threshold conditions (1, 2). The elevated morbidity and mortality associated with BD make it a major public health problem. Despite advances in recent years, the underlying neurobiology of BD remains largely unknown.

The overall heritability of BD has been estimated to be more than 70% based on twin studies (3, 4). Genome-wide association studies (GWAS) have identified several risk loci. These include markers near *ADCY2, ANK3, CACNA1C, TENM4, SYNE1, TRANK1*, and a tight cluster of genes on chromosome 3p21, among others (5-9). These loci account for only a small portion of the heritability of BD, suggesting that additional risk loci remain to be discovered.

The highly polygenic architecture of BD (10) suggests that identification of additional risk loci will require larger samples than have been studied so far. As the diagnosis of BD can be challenging, great care must be taken in the selection of cases (11). Accordingly, the ascertainment of well-characterized cases has proven to be a limiting factor. To address this problem, the present study augmented previously published case sets with a large set of well-characterized cases followed on lithium for at least 6 months and assembled by the Consortium on Lithium Genetics (ConLiGen) (12, 13). These cases were included in a recently published GWAS of lithium response (13), but have heretofore not been used for GWAS of BD itself. Since an excess of controls beyond the traditional 1:1 case:control ratio can confer a meaningful increase in power in GWAS (14-16), we have also included over twenty thousand genotyped controls drawn from public databases. Most have not, to our knowledge, been included in

previous GWAS of BD.

Surprisingly few of the published GWAS of BD have reported results for X chromosome markers, even though family and genetic linkage studies have long suggested a role for X-linked genes in BD (17-19). While the smaller effective sample size of X-linked markers necessarily leads to reduced power relative to autosomal markers (20), omission of the X-chromosome represents a considerable loss of potential association signals, since it comprises approximately 5% of the female and 2.5% of the male genome. One reason for the omission may be the relative paucity of association methods that correctly account for the added complexities of X-linked markers. Recent advances have improved the available analytic tools (21, 22), and we employ one such tool in the present study. We also employ a large X-chromosome imputation reference panel from the 1000 Genomes Project (23) that was not available during the first generation of GWAS.

In summary, the present study aimed to identify additional BD risk loci by carrying out a GWAS with new cases, excess controls, and analytic methods optimized for the X-chromosome. The most significant single nucleotide polymorphisms (SNPs) were tested for association in an independent replication sample of about 2,300 cases and 3,500 controls from two independent GWAS of BD. While we did not detect any genome-wide significant variants on chromosome X, we did find genome-wide significant evidence for common risk variants at two novel and four known autosomal loci. The results add to a growing list of common autosomal markers associated with BD and illustrate the power of well-characterized cases, combined with an excess of controls, to identify previously unknown loci involved in common, polygenic disorders.

## RESULTS

A total of 7,647 cases and 27,303 controls were analyzed in Stage 1 (Table 1), in which a total of 9,692,718 autosomal markers passed quality control. The Stage 1 studies had >90% power to detect association at the significance level of $P < 1 \times 10^{-6}$ with a common autosomal allele that confers a genotype-relative risk (GRR) of 1.15. The p-value distributions were unbiased for each of the sub-studies: all standardized genome-wide inflation factors ($\lambda_{1000}$) were $\leq 1.07$ (Figure S1). Meta-analysis of the Stage 1 studies identified 62 variants that exceeded the standard genome-wide significance threshold (Table S1). All lay within two known risk loci[7,9], one SNP (rs9834970, $P = 3.19 \times 10^{-8}$, OR = 0.88) lay near the gene *TRANK1*; all others were located in the gene *MAD1L1*. All 179 variants with fixed-effect $P < 10^{-6}$ were carried forward to the Stage 2 samples for further validation (Table S1). After LD-pruning at $r^2$=0.2, the Stage 1 results appeared to represent 14 distinct regions.

Meta-analysis of the combined Stage 1 and Stage 2 studies, comprising 9,784 cases and 30,471 controls, had >90% power to detect association at the genome-wide significance level of $P < 5.0 \times 10^{-8}$ with a common autosomal allele that confers a GRR of 1.15. The meta-analysis produced an unbiased distribution of p-values (Figures S2): The genome-wide inflation factor was 1.08, while the standardized genomic inflation factor ($\lambda_{1000}$) was ~1.01.

Six autosomal loci exceeded genome-wide significance (Figure 1). Four of these loci have been described before (Figure S3) (7-9, 24). The remaining two loci are novel BD risk loci (Figure 2). None of the six genome-wide significant loci identified here demonstrated significant heterogeneity in effect sizes across the samples studied (Table 2).

The first novel locus is located on chromosome 17q12. The most significant SNP (rs2517959, $P = 4.53 \times 10^{-9}$, OR = 1.13) is located in an intronic region of the gene, *ERBB2,* which encodes a receptor tyrosine kinase. Several other genes also lie nearby (Figure 1).

The top SNP in the second novel locus, rs12553324, lies within an intergenic region on chromosome 9p21.3 ($P = 5.87 \times 10^{-9}$, OR = 1.12). One SNP in moderate LD (rs10965780; $r^2 = 0.604$ (25)) lies within the promoter flanking region of *ELAVL2*, which encodes a neuron-specific RNA binding protein (Refseq, November 2015).

The most significant association signal in this study falls within the *MAD1L1* gene on chromosome 7p22.3, and has been reported by previous studies of BD or BD plus schizophrenia (SCZ) (8, 26). The top SNP, rs4236274, is located in an intron of *MAD1L1* ($P = 8.49 \times 10^{-12}$, OR = 0.87). An additional 60 variants at this locus surpassed the genome-wide significance threshold.

The second most significant finding in this study lies near the gene *TRANK1* on chromosome 3p22.2. This has been identified as a genome-wide significant risk locus for BD by two previous GWAS with partially overlapping samples (7, 9). The same SNP and allele of rs9834970 from those two studies was also significantly associated with BD in the present study ($P = 4.83 \times 10^{-10}$, OR = 0.88).

Twelve variants on chromosome 6q16.1 and ten variants on 12q13.1 also met the genome-wide significance threshold. The top SNP within the 6q16.1 locus (rs1487441, $P = 2.58 \times 10^{-8}$, OR = 1.12) is located in an intergenic region between *MIR2113* and *POU3F2*. Within the 12q13.1 locus, the top SNP (rs1054442, $P = 1.20 \times 10^{-8}$, OR = 1.13) is located within the 3'-UTR of *DDN*, which encodes dendrin, a cytoskeletal protein expressed at the synapse (27).

We also note nominally significant ($P$ <0.01) support for most of the loci identified in previous GWAS of BD, including markers on chromosome 3p21, and near the genes *ADCY2, ANK3, CACNA1C, LMAN2L, NCAN, TENM4,* and *SYNE1* (Table S2).

Several of the identified loci contained multiple SNPs in tight LD. To clarify whether each locus represented a single association signal, we conducted an approximate conditional analysis using GCTA (see Methods). The results suggested that each of the six GWAS-significant loci (Table 2) was consistent with only one distinct signal (Table S3).

In the X-chromosome analysis, a total of 218,707 markers passed stringent quality control. The power analysis suggested that the Stage 1 studies had 65% power to detect a common X-linked allele that confers a GRR of 1.2 at the significance level of $p<1 \times 10^{-6}$, while the combined Stage 1 and 2 meta-analysis had 66% power to detect association with the same allele at the genome-wide significance level of $p<5.0 \times 10^{-8}$. No X-chromosome markers met the $p<1 \times 10^{-6}$ threshold to be carried forward from Stage 1 to Stage 2. Consequently we did not identify any genome-wide significant signals on the X chromosome (Figure 1).

We also assessed the relative distribution of genomic heritability represented in the Stage 1 studies. Consistent with a highly polygenic architecture, there was a strong linear relationship between the genomic heritability attributable to each chromosome and chromosome length ($P$ = 0.0004, $R^2$ = 0.45) (Figure S4), except for the X-chromosome. Unlike autosomal SNPs, X-linked SNPs explained an unexpectedly small proportion of the genomic heritability of BD in this study (0.2%, Figure S4).

## DISCUSSION

This study is the first GWAS of BD to include samples from the ConLiGen study (13) and to our knowledge the largest published to date. The full set of summary results is available for download at http://www.nimh.nih.gov/hgb-data /BP_GWAS_meta_results.tar.gz. The analysis identified two novel and four known BD risk loci. The results also provided nominally significant support for most loci identified in previous GWAS of BD. No significant X-chromosome associations were detected and X-linked markers explained very little of the genomic heritability of BD.

This study has several limitations. The total sample size is still too small to identify risk variants with very small effect sizes or low frequencies, especially any residing on the X chromosome. Larger scale studies, such as those ongoing within the Psychiatric Genomics Consortium, will be needed to identify such loci. It has been shown that meaningful increases in statistical power for case-control studies can be obtained by increasing the control-to-case ratio in the range of 4 to 5 (14, 15). Thus we used an excess of controls, including unscreened controls from WTCCC2 and dbGaP. Such unscreened controls are generally easy to obtain and inexpensive, but the actual gain in power may be less than the sample size alone suggests, since some might qualify as cases if examined. However, the population prevalence of BD is probably less than 2% (1, 2), so the impact of undetected cases on power should be small (28). Another limitation arises from the fact that cases enrolled by different studies were ascertained differently, assessed with different diagnostic tools, and fulfilled different, albeit similar, diagnostic criteria. In any case, heterogeneity of ascertainment and clinical diagnosis would tend to reduce power, not lead to false positives (29, 30). This study included more than 2000 BD cases and twenty thousand

controls that have not been included in previous GWAS of BD, but most of the cases analyzed in this study have been included in previous studies. Thus, the nominally significant support we observed for many loci identified in previous GWAS cannot be considered as independent evidence of replication. The purpose of this study was not to replicate previous findings but to identify additional risk loci for BD.

This study identified two novel loci associated with BD at genome-wide significance. The top SNP within the novel BD risk locus on chromosome 17q12 lies within an intron of *ERBB2* (Erb-B2 receptor tyrosine kinase 2). Data available through the GTEx Portal (http://www.gtexportal.org/) (31) suggests that this SNP acts as an expression quantitative trait locus (eQTL) for *ERBB2* in neural tissue ($P = 2.6 \times 10^{-8}$), but other eQTLs are also present at this locus. *ERBB2* is expressed in the brain and encodes a member of the epidermal growth factor (EGF) family of receptor tyrosine kinases, which can form homo- or hetero-dimers with other ERBB proteins. Interestingly, the ERBB proteins act as cell surface receptors for neuregulins (32). This pathway has long been thought to contribute to the pathogenesis of both BD and SCZ (33-37). Two previous gene expression studies have implicated *ERBB2* in BD (38, 39), but to our knowledge the present study is the first to demonstrate genetic association with BD. If confirmed in future studies, this locus might be a promising target for novel therapeutics.

The novel risk locus identified on chromosome 9p21.3 is within an intergenic region with no known protein coding genes nearby. A total of 17 highly linked variants in this region (spanning about 17 kb) met the genome-wide significance threshold. Genomic sequence alignments from multiple species suggest that the BD-associated segment is conserved in higher primates. The top SNP is in moderate LD with another SNP that lies within the promoter flanking region of *ELAVL2*, which encodes a neuron-specific RNA binding protein that promotes neuronal

development (40). Different SNPs at this locus have been previously implicated in a GWAS of SCZ (41).

Two previous GWAS have reported suggestive evidence of association between BD and common risk variants near *MAD1L1* (8, 26). The most significant SNP (rs4332037) reported in the PGC-BD (8) study, which was not genome-wide significant, lies about 54 kb away from -- and is in low linkage disequilibrium (CEU $r^2 = 0.09$) with -- our top SNP. Ruderfer and colleagues (26) identified *MAD1L1* as a genome-wide significant locus in a GWAS that used both BD and SCZ as a combined case definition. No previous study has to our knowledge demonstrated genome-wide significant association between this locus and BD only.

The MooDS-PGC study (9), which overlaps partially with the current study, identified two novel risk loci for BD. One of them is located on 6q16.1, a region between *MIR2113* and *POU3F2*. Our top SNP in this locus (rs1487441) is in perfect LD with rs12202969, the most significant variant in the original MooDS-PGC study. Thus the present study supports the MooDS-PGC finding in a larger sample, but cannot be considered an independent replication.

Another previously reported locus that is supported by the present study is on chromosome 12q13.1. Green et al. (24) reported that an intergenic polymorphism (rs7296288) between *RHEBL1* and *DHH* was significantly associated with BD. In the present study, an SNP (rs105442) in moderate LD ($r^2 = 0.62$) with that reported by Green et al. was also significantly associated with BD. This SNP (rs105442) is located in the 3'-UTR of *DDN,* encoding dendrin. Dendrin is a synaptic protein that is markedly depleted by sleep deprivation (42), a common trigger of mood episodes in BD. Data available through the GTEx Portal (http://www.gtexportal.org/) (31) suggests that rs105442 acts as an eQTL locus for *DDN* in

skeletal muscle tissue ($P = 1.1 \times 10^{-19}$).

In this study, we also carried out an X-chromosome meta-analysis that took advantage of the latest imputation methods and the best available analysis techniques (22, 43). The available sample size was underpowered to detect a common allele that confers a GRR as low as 1.2. Larger studies are needed to rule out the involvement of common variants within this range of GRR. However, the genomic heritability analyses showed that markers on the X- chromosome contributed very little to the total genomic heritability of BD. While the SNP density on this chromosome is lower than on the autosomes, the results suggest that there may be few, if any, X-linked markers that play an important role in risk for BD. The association method used in this study is powerful under the assumption of random X-inactivation (43), but may not be optimal when X-inactivation is skewed or incomplete (44). Some candidate-gene association studies (45, 46) have reported associations between BD and genes within the pseudoautosomal regions (PARs) of chromosome X. We cannot evaluate association with the PAR, since too few markers passed QC for inclusion in the present study.

This meta-analysis study has identified two novel risk loci for BD. The findings support previous work and also suggest novel genetic influences in BD. Increasing sample sizes should enable the identification of additional risk loci for BD, but X-linked markers seem to play a smaller than expected role in this common and severe mental illness.

## MATERIALS AND METHODS

### Overall study design

A detailed description of the study design and phenotype assessments for all samples can be found in the Supplementary Materials. This study was carried out in two-stages. Stage 1 comprised a meta-analysis of directly genotyped and imputed SNPs in 7,467 patients diagnosed with BD by direct interview and 27,303 controls drawn from published BD case-control studies, dbGaP, and WTCCC2. In Stage 2, a set of 179 variants from Stage 1 with $P < 1.0 \times 10^{-6}$ were tested in an independent sample of 2,313 BD cases and 3,486 controls from two ongoing Swedish GWAS.

### Stage 1 samples

Stage 1 comprised cases and controls from five published BD GWAS studies (Table 1) and an independent set of 2,072 European-ancestry patients with a long-term diagnosis of BD who were treated with lithium and submitted to the Consortium on Lithium Genetics (ConLiGen) (12). The ConLiGen sample has not been used in any previous GWAS studies of BD. An additional 20,998 genotyped individuals obtained from dbGaP and the WTCCC2 were used as controls. All participants included in the final association tests were of European ancestry.

### Stage 2 samples

The Stage 2 sample was predominantly Swedish and exclusively Northern European. BD cases from Sweden were collected through two recruitment streams: 1,908 cases from the Stanley study (47, 48) and 229 cases from the St. Göran Bipolar Project (49). Most control subjects

(n=3,113) were randomly selected from Swedish population registers (ascertained on a national basis) and 55 were from the St. Göran Bipolar Project (49). The exclusion criterion for controls was any hospitalization for SCZ or BD. DNA collection procedures have been previously described (50). Both projects were approved by the Regional Ethical Review Board in Stockholm (Sweden), and all participants provided written informed consent. Genotyping was conducted at the Broad Institute of Harvard and MIT using Affymetrix 6.0 (Swedish sample 1) and Illumina OmniExpress (Swedish sample 2) chips.

## Quality control

Quality control procedures were carried out separately in each data set. The quality control parameters for retaining SNPs and subjects were: Subject missingness < 0.03; autosomal heterozygosity rate within mean ± 3SD; minor allele frequency (MAF) >= 0.01 (for Affymetrix data, we kept all variants with a MAF >= 0.05); SNP missingness < 0.05; and SNP Hardy-Weinberg equilibrium $P > 10^{-6}$ in controls (for markers on the X chromosome, only females were used for the Hardy-Weinberg equilibrium tests), no discrepancies between reported sex and sex determined by genotypes on chromosome X. For each data set, around 100K LD-pruned SNPs were used to identify duplicated samples, cryptically related subjects, and population outliers. Relatedness testing was carried out by PLINK. Duplicated samples and cryptically related pairs were identified (Pi_hat > 0.10); only one member of each pair was kept for the data analyses. To identify cryptically related subjects present across different data sets, we tested approximately 20K LD-pruned SNPs present in all of the SNP arrays used by any of the studies. EIGENSOFT (51) was used to identify population outliers (Figure S5). All subjects with apparent non-European ancestry were excluded from the data analyses. After basic QC, we

matched data sets by genotyping platform (Table 1).

## Imputation

Genotype imputation was performed using the prephasing/imputation strategy (52) by SHAPEIT2 (53) and IMPUTE2 (54) for both autosomes and the X chromosome. Phase 3 of the 1000 Genomes Project data set (n=2,504) was used as the reference panel for imputation. Appropriate options were used for X chromosome phasing and imputation (--chrX was set for SHAPEIT2 for phasing, and --chrX and gender information were provided to IMPUTE2 for imputation).

## Genotype–phenotype association analysis

Gene dosages for all markers with an imputation INFO $\geq$ 0.5 were used for the final association tests. In total, over 9M genotyped or imputed autosomal SNPs were analyzed. Genotype–phenotype association with imputed allele dosages for autosomal SNPs was carried out with logistic regression as implemented by PLINK (55). The first 10 principal components of population structure were used as covariates in the analyses.

Association tests for markers on the X chromosome were performed with Clayton's "one degree-of-freedom" test, implemented in the snpStats R package (22, 56). An extensive simulation study (43) of several different tests designed specifically for chromosome X association testing concluded that Clayton's "one degree-of-freedom" statistic is robust and powerful across a wide range of realistic conditions.

## Meta-analysis of Stage 1 samples

Meta-analysis in Stage 1 was conducted using the sample size-based method in METAL (57). Meta-analysis results were corrected with genomic control to eliminate any residual bias.

## Selection of SNPs for Stage 2

All markers with a p-value $\leq 1.0 \times 10^{-6}$ in Stage 1 (n=179) were selected for validation in the Stage 2 samples, using identical QC and genotype-phenotype association procedures. Of these SNPs, results were available in the Stage 2 samples for 144 SNPs.

## Meta-analysis of Stage 1 and Stage 2 samples

Meta-analysis of the Stage 1 and 2 results was conducted using the sample size-based method in METAL, under a fixed effects model. Again, genomic control was used to eliminate any residual bias. Association results were considered genome-wide significant if the meta p-values were $< 5 \times 10^{-8}$ and the heterogeneity tests were not significant ($p>0.05$).

## Identification of distinct association signals within risk loci

GCTA (58) was used to identify independently associated variants within each of the 6 GWAS-significant loci. GCTA made use of (i) summary statistics from the meta-analysis of the stage 1 and stage 2 samples, and (ii) genotype data from a reference sample for LD estimation between markers. Study-5 (416 BP cases and 2,741 controls) was used as the reference sample here because subjects from Study-5 were genotyped by the highest density SNP array (Illumina HumanOmni2.5M). GCTA uses a stepwise selection strategy (59) to identify any independent signals through an approximate conditional association analysis.

## Genomic heritability estimation of Stage 1 samples

We estimated the genomic heritability (the variance in case-control status explained by all genotyped markers) using a linear mixed model developed by Lee et al. (60). Briefly, a genome relationship matrix (GRM) between all pairs of individuals was estimated from the genome-wide markers. The GRM was then used to estimate variance attributable to markers via residual maximum likelihood (REML) analysis with the first 10 principal components of population structure used as covariates. The results were transformed to the liability scale by assuming a disease prevalence of 2% for BD. For the genomic heritability estimation, we only retained autosomal markers with MAF >0.01 and imputation INFO ≥ 0.9.

We partitioned genomic heritability attributable to each autosome by estimating the GRM from SNPs on each autosome and then fitting the GRMs of all the autosomes simultaneously in the linear mixed model. We further estimated the variance attributable to the X chromosome using a method developed by Yang and colleagues (58) under the assumption of full dosage compensation (complete inactivation of one X chromosome for females).

## Power estimation

Power analysis of autosomal markers was done with the Genetic Power Calculator (GPC) (61) under the following assumptions: Trait prevalence 2%, risk allele and marker allele frequency 25%, D-prime 1, GRR of 1.15 under a log-additive model. Since excess controls were used in this study and case/control ratio varied across sub-studies, simply using the total sample size might overestimate the power. To take this into account, we used GPC to estimate the non-centrality parameter (NCP) for each study and then iteratively determined the effective symmetric case/control sample size that returns the same NCP (62). The total effective sample sizes of the Stage 1 studies and Stage 1+2 studies were then used for the power estimation.

Power analysis for markers on chromosome X was done with XGWAS (https://github.com/PeteHaitch/XGWAS), which uses a simulation-based method to estimate power. The total effective sample sizes of Stage 1 studies and Stage 1+2 were used for simulations.

**FUNDING**

This work was supported by the National Institute of Mental Health (NIMH) Intramural Research Program (ZIA-MH00284311; NCT00001174) and a NARSAD Young Investigator Award to L.H.

**ACKNOWLEDGEMENTS**

This work utilized the computational resources of the NIH HPC Biowulf cluster (http://hpc.nih.gov)

We obtained several datasets from the database of Genotypes and Phenotypes. A list of investigators and the funding support for these datasets can be found by searching http://www.ncbi.nlm.nih.gov/gap using these accession numbers: phs000124.v2.p1, phs000404.v1.p1, phs000303.v1.p1, phs000203.v1.p1, phs000170.v1.p1, phs000188.v1.p1, and phs000237.v1.p1

This study makes use of data generated by the Wellcome Trust Case-Control Consortium 2 (EGAD00000000021 and EGAD00000000023). We thank the Wellcome Trust Case Control Consortium for making data available for further analysis.

Some data and biomaterials were provided by the Rutgers Cell and DNA Repository.

Some data and biomaterials were collected in four projects that participated in the National Institute of Mental Health (NIMH) Bipolar Disorder Genetics Initiative. From 1991-98, the Principal Investigators and Co-Investigators were: Indiana University, Indianapolis, IN, U01 MH46282, John Nurnberger, M.D., Ph.D., Marvin Miller, M.D., Howard J. Edenberg, Ph.D. and Elizabeth Bowman, M.D.; Washington University, St. Louis, MO, U01 MH46280, Theodore Reich, M.D., Allison Goate, Ph.D., and John Rice, Ph.D.; Johns Hopkins University, Baltimore, MD, U01 MH46274, J. Raymond DePaulo, Jr., M.D., Sylvia Simpson, M.D., MPH, and Colin Stine, Ph.D.; NIMH Intramural Research Program, Clinical Neurogenetics Branch, Bethesda, MD, Elliot Gershon, M.D., Diane Kazuba, B.A., and Elizabeth Maxwell, M.S.W.

NIMH Study 1 -- Data and biomaterials were collected as part of ten projects that participated in the National Institute of Mental Health (NIMH) Bipolar Disorder Genetics Initiative. From 1999-03, the Principal Investigators and Co-Investigators were: Indiana University, Indianapolis, IN, R01 MH59545, John Nurnberger, M.D., Ph.D., Marvin J. Miller, M.D., Elizabeth S. Bowman, M.D., N. Leela Rau, M.D., P. Ryan Moe, M.D., Nalini Samavedy, M.D., Rif El-Mallakh, M.D.

(at University of Louisville), Husseini Manji, M.D. (at Wayne State University), Debra A. Glitz, M.D. (at Wayne State University), Eric T. Meyer, M.S., Carrie Smiley, R.N., Tatiana Foroud, Ph.D., Leah Flury, M.S., Danielle M. Dick, Ph.D., Howard J. Edenberg, Ph.D.; Washington University, St. Louis, MO, R01 MH059534, John Rice, Ph.D, Theodore Reich, M.D., Allison Goate, Ph.D., Laura Bierut, M.D.; Johns Hopkins University, Baltimore, MD, R01 MH59533, Melvin McInnis, M.D., J. Raymond DePaulo, Jr., M.D., Dean F. MacKinnon, M.D., Francis M. Mondimore, M.D., James B. Potash, M.D., Peter P. Zandi, Ph.D, Dimitrios Avramopoulos, and Jennifer Payne; University of Pennsylvania, PA, R01 MH59553, Wade Berrettini, M.D., Ph.D.; University of California at Irvine, CA, R01 MH60068, William Byerley, M.D., and Mark Vawter, M.D.; University of Iowa, IA, R01 MH059548, William Coryell, M.D., and Raymond Crowe, M.D.; University of Chicago, IL, R01 MH59535, Elliot Gershon, M.D., Judith Badner, Ph.D., Francis McMahon, M.D., Chunyu Liu, Ph.D., Alan Sanders, M.D., Maria Caserta, Steven Dinwiddie, M.D., Tu Nguyen, Donna Harakal; University of California at San Diego, CA, R01 MH59567, John Kelsoe, M.D., Rebecca McKinney, B.A.; Rush University, IL, R01 MH059556, William Scheftner, M.D., Howard M. Kravitz, D.O., M.P.H., Diana Marta, B.S., Annette Vaughn-Brown, M.S.N., R.N., and Laurie Bederow, M.A.; NIMH Intramural Research Program, Bethesda, MD, 1Z01MH002810-01, Francis J. McMahon, M.D., Layla Kassem, PsyD, Sevilla Detera-Wadleigh, Ph.D, Lisa Austin, Ph.D, Dennis L. Murphy, M.D.

NIMH Study 19 -- Data and biomaterials were collected for the Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD), a multi-center, longitudinal (5-8 year) project selected from responses to RFP #NIMH-98-DS-0001, "Treatment for Bipolar Disorder." The project was led by Gary Sachs, M.D., and coordinated by Massachusetts General Hospital in Boston, MA. The NIMH grant number was 2N01MH080001-001.

NIMH Study 40 -- Data and biomaterials were collected as part of eleven projects (Study 40) that participated in the National Institute of Mental Health (NIMH) Bipolar Disorder Genetics Initiative. From 2003-2007, the Principal Investigators and Co-Investigators were: Indiana University, Indianapolis, IN, R01 MH59545, John Nurnberger, M.D., Ph.D., Marvin J. Miller, M.D., Elizabeth S. Bowman, M.D., N. Leela Rau, M.D., P. Ryan Moe, M.D., Nalini Samavedy, M.D., Rif El-Mallakh, M.D. (at University of Louisville), Husseini Manji, M.D. (at Johnson and Johnson), Debra A. Glitz, M.D. (at Wayne State University), Eric T. Meyer, Ph.D., M.S. (at Oxford University, UK), Carrie Smiley, R.N., Tatiana Foroud, Ph.D., Leah Flury, M.S., Danielle M. Dick, Ph.D (at Virginia Commonwealth University), Howard J. Edenberg, Ph.D.; Washington University, St. Louis, MO, R01 MH059534, John Rice, Ph.D, Theodore Reich, M.D., Allison Goate, Ph.D., Laura Bierut, M.D. K02 DA21237; Johns Hopkins University, Baltimore, M.D., R01 MH59533, Melvin McInnis, M.D., J. Raymond DePaulo, Jr., M.D., Dean F. MacKinnon, M.D., Francis M. Mondimore, M.D., James B. Potash, M.D., Peter P. Zandi, Ph.D, Dimitrios Avramopoulos, and Jennifer Payne; University of Pennsylvania, PA, R01 MH59553, Wade Berrettini, M.D., Ph.D.; University of California at San Francisco, CA, R01 MH60068, William Byerley, M.D., and Sophia Vinogradov, M.D.; University of Iowa, IA, R01 MH059548, William Coryell, M.D., and Raymond Crowe, M.D.; University of Chicago, IL, R01 MH59535, Elliot Gershon, M.D., Judith Badner, Ph.D., Francis McMahon, M.D., Chunyu Liu, Ph.D., Alan Sanders, M.D., Maria Caserta, Steven Dinwiddie, M.D., Tu Nguyen, Donna Harakal; University of California at San Diego, CA, R01 MH59567, John Kelsoe, M.D., Rebecca McKinney, B.A.; Rush University, IL, R01 MH059556, William Scheftner, M.D.,

Howard M. Kravitz, D.O., M.P.H., Diana Marta, B.S., Annette Vaughn-Brown, M.S.N., R.N., and Laurie Bederow, M.A.; NIMH Intramural Research Program, Bethesda, MD, 1Z01MH002810-01, Francis J. McMahon, M.D., Layla Kassem, Psy.D., Sevilla Detera-Wadleigh, Ph.D, Lisa Austin, Ph.D, Dennis L. Murphy, M.D.; Howard University, William B. Lawson, M.D., Ph.D., Evarista Nwulia, M.D., and Maria Hipolito, M.D. This work was supported by the NIH grants P50CA89392 from the National Cancer Institute and 5K02DA021237 from the National Institute of Drug Abuse.

The Swedish samples were funded by the Stanley Center for Psychiatric Research, Broad Institute from a grant from Stanley Medical Research Institute, the Swedish Research Council (K2014-62X-21445-05-3, K2012-63X-21445-03-2, K2014-62X-14647-12-51 and K2010-61P-21568-01-4), the Söderström-Königska Foundation (SLS-472751), the Swedish foundation for Strategic Research (KF10-0039), and the China Scholarship Council. We would like to thank all of the participants who have kindly given their time and DNA to participate in our research, the Swedish quality register for bipolar disorders (BipoläR), and data collectors at the Department of Medical Epidemiology and Biostatistics (MEB) at Karolinska Institutet for help with recruitment of participants.

The ConLiGen project was in part funded by the Deutsche Forschungsgemeinschaft (DFG; grant no. RI 908/7-1) and the Intramural Research Program of the National Institute of Mental Health (ZIA- MH00284311; NCT00001174). The Romanian sample included in the ConLiGen project was also funded by UEFISCDI, Romania (grant no. 89/2012 awarded to Maria Grigoroiu-Serbanescu, PhD) and BMBF, Germany, grant no. 01GS08144 awarded to M. M. Nöthen and S. Cichon). The Canadian sample has been part of a study funded by the Canadian Institutes of Health Research (grant 64410 to MA). The collection of the Barcelona sample was supported by the Centro de Investigación en Red de Salud Mental (CIBERSAM), IDIBAPS (grant numbers PI080247, PI1200906, PI12/00018) and Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement (2014SGR1636 and 2014SGR398).

Genotyping was funded in part by the German Federal Ministry of 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 Programme (grant 01ZX1314G to awarded to Thomas G. Schulze and Marcella Rietschel, grant 01ZX1314A to Sven Cichon and Markus M. Nöthen).

Thomas G. Schulze and Urs Heilbronner received support from the Dr.-Lisa-Oehler-Foundation (Kassel, Germany). Andrea Pfennig, Thomas Stamm, Michael Bauer, Andreas Reif, and Thomas G. Schulze received support from the German Federal Ministry of Education and Research (BMBF) within the framework of the *BipolLife* network (www.bipolife.org).

M.M.N. is a member of the DFG-funded Excellence-Cluster ImmunoSensation. M.M.N. also received support from the Alfried Krupp von Bohlen und Halbach-Stiftung. The study was supported by the German Research Foundation (DFG; grant ConLiGen RI 908/7-1; grant SFB 636 Z4; grant FOR2107, RI 908/11-1 to M.R., WI 3439/3-1 to S.H.W., NO 246/10-1 to M.M.N.).

The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health. Additional funds were provided by the NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. Donors were enrolled at Biospecimen Source Sites funded by NCI\SAIC-Frederick, Inc. (SAIC-F) subcontracts to the National Disease Research Interchange (10XS170), Roswell Park Cancer Institute (10XS171), and Science Care, Inc. (X10S172). The Laboratory, Data Analysis, and Coordinating Center (LDACC) was funded through a contract (HHSN268201000029C) to The Broad Institute, Inc. Biorepository operations were funded through an SAIC-F subcontract to Van Andel Institute (10ST1035). Additional data repository and project management were provided by SAIC-F (HHSN261200800001E). The Brain Bank was supported by a supplements to University of Miami grants DA006227 & DA033684 and to contract N01MH000028. Statistical Methods development grants were made to the University of Geneva (MH090941 & MH101814), the University of Chicago (MH090951, MH090937, MH101820, MH101825), the University of North Carolina - Chapel Hill (MH090936 & MH101819), Harvard University (MH090948), Stanford University (MH101782), Washington University St Louis (MH101810), and the University of Pennsylvania (MH101822). The data used for the analyses described in this manuscript were obtained from the GTEx Portal on 10/12/2015.

The Australian cohort collection was supported by the Australian National Health and Medical Research Council (NHMRC) Program Grants 510135 (PBM) and 1037196 (PBM & PRS); and Project Grants 1063960 (JMF & PRS) and 1066177 (JMF).

The Swiss samples were funded by the Swiss National Science Foundation (grant number 32003B_125469 and NCCR Synapsy, Jean-Michel Aubry aand Alexandre Dayer).

Most importantly, we thank the individuals and families who have participated in and contributed time and data to these studies.

## REFERENCES


1 Merikangas, K.R., Jin, R., He, J.P., Kessler, R.C., Lee, S., Sampson, N.A., Viana, M.C., Andrade, L.H., Hu, C., Karam, E.G. *et al.* (2011) Prevalence and correlates of bipolar spectrum disorder in the world mental health survey initiative. *Arch. Gen. Psychiatry*, **68**, 241-251.

2 Merikangas, K.R., Akiskal, H.S., Angst, J., Greenberg, P.E., Hirschfeld, R.M., Petukhova, M. and Kessler, R.C. (2007) Lifetime and 12-month prevalence of bipolar spectrum disorder in the National Comorbidity Survey replication. *Arch. Gen. Psychiatry*, **64**, 543-552.

3 McGuffin, P., Rijsdijk, F., Andrew, M., Sham, P., Katz, R. and Cardno, A. (2003) The heritability of bipolar affective disorder and the genetic relationship to unipolar depression. *Arch. Gen. Psychiatry*, **60**, 497-502.

4 Edvardsen, J., Torgersen, S., Roysamb, E., Lygren, S., Skre, I., Onstad, S. and Oien, P.A. (2008) Heritability of bipolar spectrum disorders. Unity or heterogeneity? *J. Affect. Disord.*, **106**, 229-240.

5 Baum, A.E., Akula, N., Cabanero, M., Cardona, I., Corona, W., Klemens, B., Schulze, T.G., Cichon, S., Rietschel, M., Nothen, M.M. *et al.* (2008) A genome-wide association study implicates diacylglycerol kinase eta (DGKH) and several other genes in the etiology of bipolar disorder. *Mol. Psychiatry*, **13**, 197-207.

6 Ferreira, M.A., O'Donovan, M.C., Meng, Y.A., Jones, I.R., Ruderfer, D.M., Jones, L., Fan, J., Kirov, G., Perlis, R.H., Green, E.K. *et al.* (2008) Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. *Nat. Genet.*, **40**, 1056-1058.

7 Chen, D.T., Jiang, X., Akula, N., Shugart, Y.Y., Wendland, J.R., Steele, C.J., Kassem, L., Park, J.H., Chatterjee, N., Jamain, S. *et al.* (2013) 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.

8 Psychiatric GWAS Consortium Bipolar Disorder Working Group. (2011) Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat. Genet.*, **43**, 977-983.

9 Muhleisen, T.W., Leber, M., Schulze, T.G., Strohmaier, J., Degenhardt, F., Treutlein, J., Mattheisen, M., Forstner, A.J., Schumacher, J., Breuer, R. *et al.* (2014) Genome-wide association study reveals two new risk loci for bipolar disorder. *Nat Commun*, **5**, 3339.

10 Moser, G., Lee, S.H., Hayes, B.J., Goddard, M.E., Wray, N.R. and Visscher, P.M. (2015) Simultaneous discovery, estimation and prediction analysis of complex traits using a bayesian mixture model. *PLoS Genet*, **11**, e1004969.

11 Phillips, M.L. and Kupfer, D.J. (2013) Bipolar disorder diagnosis: challenges and future directions. *Lancet*, **381**, 1663-1671.

12 Schulze, T.G., Alda, M., Adli, M., Akula, N., Ardau, R., Bui, E.T., Chillotti, C., Cichon, S., Czerski, P., Del Zompo, M. *et al.* (2010) The International Consortium on Lithium Genetics (ConLiGen): an initiative by the NIMH and IGSLI to study the genetic basis of response to lithium treatment. *Neuropsychobiology*, **62**, 72-78.

13 Hou, L., Heilbronner, U., Degenhardt, F., Adli, M., Akiyama, K., Akula, N., Ardau, R., Arias, B., Backlund, L., Banzato, C.E. *et al.* (2016) Genetic variants associated with response to lithium treatment in bipolar disorder: a genome-wide association study. *Lancet*, **387**, 1085-1093.

14 Hennessy, S., Bilker, W.B., Berlin, J.A. and Strom, B.L. (1999) Factors influencing the optimal control-to-case ratio in matched case-control studies. *Am. J. Epidemiol.*, **149**, 195-197.

15 Kang, M., Choi, S. and Koh, I. (2009) The Effect of Increasing Control-to-case Ratio on Statistical Power in a Simulated Case-control SNP Association Study. *Genomics Inform*, **7**, 148-151.
16 Hong, E.P. and Park, J.W. (2012) Sample size and statistical power calculation in genetic association studies. *Genomics Inform*, **10**, 117-122.
17 Rosanoff, A.J., Handy, L. and Plesset, I.R. (1935) The etiology of manic- depressive syndromes with special reference to their occur- rence in twins. *Am. J. Psychiatry*, 91, 725-762.
18 Reich, T., Clayton, P.J. and Winokur, G. (1969) Family history studies: V. The genetics of mania. *Am. J. Psychiatry*, **125**, 1358-1369.
19 Baron, M. (1991) X-linkage and manic-depressive illness: a reassessment. *Soc. Biol.*, **38**, 179-188.
20 Wise, A.L., Gyi, L. and Manolio, T.A. (2013) eXclusion: toward integrating the X chromosome in genome-wide association analyses. *Am. J. Hum. Genet.*, **92**, 643-647.
21 Zheng, G., Joo, J., Zhang, C. and Geller, N.L. (2007) Testing association for markers on the X chromosome. *Genet. Epidemiol.*, **31**, 834-843.
22 Clayton, D. (2008) Testing for association on the X chromosome. *Biostatistics*, **9**, 593-600.
23 1000 Genomes Project Consortium. (2015) A global reference for human genetic variation. *Nature*, **526**, 68-74.
24 Green, E.K., Hamshere, M., Forty, L., Gordon-Smith, K., Fraser, C., Russell, E., Grozeva, D., Kirov, G., Holmans, P., Moran, J.L. *et al.* (2013) Replication of bipolar disorder susceptibility alleles and identification of two novel genome-wide significant associations in a new bipolar disorder case-control sample. *Mol. Psychiatry*, **18**, 1302-1307.
25 Johnson, A.D., Handsaker, R.E., Pulit, S.L., Nizzari, M.M., O'Donnell, C.J. and de Bakker, P.I. (2008) SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics*, **24**, 2938-2939.
26 Ruderfer, D.M., Fanous, A.H., Ripke, S., McQuillin, A., Amdur, R.L., Schizophrenia Working Group of Psychiatric Genomics Consortium, Bipolar Disorder Working Group of Psychiatric Genomics Consortium, Cross-Disorder Working Group of Psychiatric Genomics Consortium, Gejman, P.V., O'Donovan, M.C. *et al.* (2014) Polygenic dissection of diagnosis and clinical dimensions of bipolar disorder and schizophrenia. *Mol. Psychiatry*, **19**, 1017-1024.
27 Kremerskothen, J., Kindler, S., Finger, I., Veltel, S. and Barnekow, A. (2006) Postsynaptic recruitment of Dendrin depends on both dendritic mRNA transport and synaptic anchoring. *J. Neurochem.*, **96**, 1659-1666.
28 Moskvina, V., Holmans, P., Schmidt, K.M. and Craddock, N. (2005) Design of case-controls studies with unscreened controls. *Ann. Hum. Genet.*, **69**, 566-576.
29 Manchia, M., Cullis, J., Turecki, G., Rouleau, G.A., Uher, R. and Alda, M. (2013) The impact of phenotypic and genetic heterogeneity on results of genome wide association studies of complex diseases. *PLoS One*, **8**, e76295.
30 Zheng, G. and Tian, X. (2005) The impact of diagnostic error on testing genetic association in case-control studies. *Stat. Med.*, **24**, 869-882.
31 GTEx Consortium. (2015) The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*, **348**, 648-660.
32 Lin, W., Sanchez, H.B., Deerinck, T., Morris, J.K., Ellisman, M. and Lee, K.F. (2000) Aberrant development of motor axons and neuromuscular synapses in erbB2-deficient mice. *Proc. Natl. Acad. Sci. U. S. A.*, **97**, 1299-1304.

33 Mei, L. and Nave, K.A. (2014) Neuregulin-ERBB signaling in the nervous system and neuropsychiatric diseases. *Neuron*, **83**, 27-49.
34 Prata, D.P., Breen, G., Osborne, S., Munro, J., St Clair, D. and Collier, D.A. (2009) An association study of the neuregulin 1 gene, bipolar affective disorder and psychosis. *Psychiatr. Genet.*, **19**, 113-116.
35 Georgieva, L., Dimitrova, A., Ivanov, D., Nikolov, I., Williams, N.M., Grozeva, D., Zaharieva, I., Toncheva, D., Owen, M.J., Kirov, G. *et al.* (2008) Support for neuregulin 1 as a susceptibility gene for bipolar disorder and schizophrenia. *Biol. Psychiatry*, **64**, 419-427.
36 Thomson, P.A., Christoforou, A., Morris, S.W., Adie, E., Pickard, B.S., Porteous, D.J., Muir, W.J., Blackwood, D.H. and Evans, K.L. (2007) Association of Neuregulin 1 with schizophrenia and bipolar disorder in a second cohort from the Scottish population. *Mol. Psychiatry*, **12**, 94-104.
37 Sei, Y., Ren-Patterson, R., Li, Z., Tunbridge, E.M., Egan, M.F., Kolachana, B.S. and Weinberger, D.R. (2007) Neuregulin1-induced cell migration is impaired in schizophrenia: association with neuregulin1 and catechol-o-methyltransferase gene polymorphisms. *Mol. Psychiatry*, **12**, 946-957.
38 Shao, L. and Vawter, M.P. (2008) Shared gene expression alterations in schizophrenia and bipolar disorder. *Biol. Psychiatry*, **64**, 89-97.
39 Bezchlibnyk, Y.B., Wang, J.F., McQueen, G.M. and Young, L.T. (2001) Gene expression differences in bipolar disorder revealed by cDNA array analysis of post-mortem frontal cortex. *J. Neurochem.*, **79**, 826-834.
40 Akamatsu, W., Okano, H.J., Osumi, N., Inoue, T., Nakamura, S., Sakakibara, S., Miura, M., Matsuo, N., Darnell, R.B. and Okano, H. (1999) Mammalian ELAV-like neuronal RNA-binding proteins HuB and HuC promote neuronal development in both the central and the peripheral nervous systems. *Proc. Natl. Acad. Sci. U. S. A.*, **96**, 9885-9890.
41 Yamada, K., Iwayama, Y., Hattori, E., Iwamoto, K., Toyota, T., Ohnishi, T., Ohba, H., Maekawa, M., Kato, T. and Yoshikawa, T. (2011) Genome-wide association study of schizophrenia in Japanese population. *PLoS One*, **6**, e20468.
42 Neuner-Jehle, M., Denizot, J.P., Borbely, A.A. and Mallet, J. (1996) Characterization and sleep deprivation-induced expression modulation of dendrin, a novel dendritic protein in rat brain neurons. *J. Neurosci. Res.*, **46**, 138-151.
43 Hickey, P.F. and Bahlo, M. (2011) X chromosome association testing in genome wide association studies. *Genet. Epidemiol.*, **35**, 664-670.
44 Wang, J., Yu, R. and Shete, S. (2014) X-chromosome genetic association test accounting for X-inactivation, skewed X-inactivation, and escape from X-inactivation. *Genet. Epidemiol.*, **38**, 483-493.
45 Etain, B., Dumaine, A., Bellivier, F., Pagan, C., Francelle, L., Goubran-Botros, H., Moreno, S., Deshommes, J., Moustafa, K., Le Dudal, K. *et al.* (2012) Genetic and functional abnormalities of the melatonin biosynthesis pathway in patients with bipolar disorder. *Hum. Mol. Genet.*, **21**, 4030-4037.
46 Geoffroy, P.A., Boudebesse, C., Henrion, A., Jamain, S., Henry, C., Leboyer, M., Bellivier, F. and Etain, B. (2014) An ASMT variant associated with bipolar disorder influences sleep and circadian rhythms: a pilot study. *Genes Brain Behav*, **13**, 299-304.
47 Karanti, A., Bobeck, C., Osterman, M., Kardell, M., Tidemalm, D., Runeson, B., Lichtenstein, P. and Landen, M. (2015) Gender differences in the treatment of patients with bipolar disorder: a study of 7354 patients. *J. Affect. Disord.*, **174**, 303-309.

48 Tidemalm, D., Haglund, A., Karanti, A., Landen, M. and Runeson, B. (2014) Attempted suicide in bipolar disorder: risk factors in a cohort of 6086 patients. *PLoS One*, **9**, e94097.
49 Jakobsson, J., Zetterberg, H., Blennow, K., Johan Ekman, C., Johansson, A.G. and Landen, M. (2013) Altered concentrations of amyloid precursor protein metabolites in the cerebrospinal fluid of patients with bipolar disorder. *Neuropsychopharmacology*, **38**, 664-672.
50 Ripke, S., O'Dushlaine, C., Chambert, K., Moran, J.L., Kahler, A.K., Akterin, S., Bergen, S.E., Collins, A.L., Crowley, J.J., Fromer, M. *et al.* (2013) Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat. Genet.*, **45**, 1150-1159.
51 Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. and Reich, D. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.*, **38**, 904-909.
52 Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. and Abecasis, G.R. (2012) Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.*, **44**, 955-959.
53 Delaneau, O., Zagury, J.F. and Marchini, J. (2013) Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods*, **10**, 5-6.
54 Howie, B.N., Donnelly, P. and Marchini, J. (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*, **5**, e1000529.
55 Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559-575.
56 Clayton, D. (2014) snpStats: SnpMatrix and XSnpMatrix classes and methods. R package version 1.18.0.
57 Willer, C.J., Li, Y. and Abecasis, G.R. (2010) METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, **26**, 2190-2191.
58 Yang, J., Lee, S.H., Goddard, M.E. and Visscher, P.M. (2011) GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.*, **88**, 76-82.
59 Yang, J., Ferreira, T., Morris, A.P., Medland, S.E., Genetic Investigation of, A.T.C., Replication, D.I.G., Meta-analysis, C., Madden, P.A., Heath, A.C., Martin, N.G. *et al.* (2012) Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.*, **44**, 369-375.
60 Lee, S.H., Wray, N.R., Goddard, M.E. and Visscher, P.M. (2011) Estimating missing heritability for disease from genome-wide association studies. *Am. J. Hum. Genet.*, **88**, 294-305.
61 Purcell, S., Cherny, S.S. and Sham, P.C. (2003) Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics*, **19**, 149-150.
62 de Bakker, P.I., Ferreira, M.A., Jia, X., Neale, B.M., Raychaudhuri, S. and Voight, B.F. (2008) Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum. Mol. Genet.*, **17**, R122-128.

**Figure legends**

**Figure 1 Manhattan plot and regional association plots.** (a) Manhattan plot for all analyzed markers (the inset is for markers on the X chromosome); (b,c) regional association plots for the two novel risk loci (left: 9p21.3, right: *ERBB2* region).

**Figure 2 Genetic effect sizes for the novel risk loci**. Forest plots displaying the odds ratios (OR) and 95% confidence intervals for the most significant SNPs in the (a) 9p21.3 region and (b) the *ERBB2* region. Horizontal lines indicate the 95% confidence interval of the OR for each study, with a shaded box around the point estimate, drawn proportional to the sample size. The diamond indicates the overall weighted OR for all samples included in the meta-analysis.

**Table 1. Study sample characteristics**

| Study | Female (%) | Platform | Samples | Cases | Controls | Total |
|---|---|---|---|---|---|---|
| | | | *Stage 1* | | | |
| Study-1 | 49.2 | Affymetrix 6.0 | GAIN-NIMH | 974 | 1,032 | 2,006 |
| Study-2 | 54.7 | Affymetrix 500K | WTCCC1-BD | 1,844 | 2,932 | 4,776 |
| | | | STEP-BD | 945 | 650 | 1,595 |
| Study-3 | 52.2 | Affymetrix 6.0 | TGEN1 | 1,184 | 400 | 1,584 |
| | | | ConLiGen | 32 | 0 | 32 |
| | | | WTCCC2 | 0 | 2724 | 2,724 |
| Study-4 | 48.7 | HumanHap550 | BoMA | 628 | 1,291 | 1,919 |
| | | | dbGaP[a] | 0 | 1,462 | 1,462 |
| Study-5 | 53.7 | HumanOmni2.5M | ConLiGen | 416 | 0 | 416 |
| | | | dbGaP[a] | 0 | 2,741 | 2,741 |
| Study-6 | 50.0 | Illumina610/660/1M/OE | ConLiGen | 1,624 | 0 | 1,624 |
| | | | dbGaP[a] | 0 | 14,071 | 14,071 |
| | | | *Stage 2* | | | |
| Swedish sample 1 | 52.2 | Affymetrix 6.0 | sw34 | 796 | 1,986 | 2,782 |
| Swedish sample 2 | 57.6 | OmniExpress | sw6 | 1,341 | 1,182 | 2,523 |
| Grand total | 51.7 | | | 9,784 | 30,471 | 40,255 |

[a] Details on contributing samples from dbGaP are given in Table S4.

**Table 2. Top SNPs in regions showing genome-wide significant evidence for association with BD in the combined analysis**

| Chr | Position[a] | SNP | Alleles[b] | FRQ[c] | P-value | OR (CI) | Directions[d] | Hetero-geneity P-value[e] |
|---|---|---|---|---|---|---|---|---|
| 3 | 36856030 | rs9834970 | T/C | 0.50 | 4.83E-10 | 0.88 (0.85 - 0.92) | - - - - - - - - | 0.15 |
| 6 | 98553894 | rs1487441 | A/G | 0.49 | 2.58E-08 | 1.12 (1.08 - 1.15) | ++++++++ | 0.78 |
| 7 | 1896413 | rs4236274 | G/A | 0.39 | 8.49E-12 | 0.87 (0.83 - 0.91) | - - - - - - - - | 0.44 |
| 9 | 23347865 | rs12553324 | G/C | 0.41 | 5.87E-09 | 1.12 (1.07 - 1.16) | +++++++ - | 0.07 |
| 12 | 49389320 | rs1054442 | C/A | 0.38 | 1.20E-08 | 1.13 (1.09 - 1.17) | ++++++++ | 0.56 |
| 17 | 37846512 | rs2517959 | A/T | 0.33 | 4.53E-09 | 1.13 (1.09 - 1.17) | ++++++++ | 0.61 |

Abbreviations: OR, odds ratio relative to major allele; CI: confidence interval (95%)

[a] Base pair position on Build 37/hg19

[b] Alleles: minor allele/major allele

[c] FRQ: Observed allele frequency of the minor allele

[d] Summary of effect direction for each sub-study

[e] P-value for the meta-analysis heterogeneity test

**ABBREVIATIONS**

BD: Bipolar Disorder

SCZ: Schizophrenia

GWAS: Genome-wide Association Study

SNP: Single Nucleotide Polymorphism

MAF: Minor Allele Frequency

OR: Odds Ratio

SD: Standard Deviation

CI: Confidence Interval

eQTL: expression Quantitative Trait Locus

GAIN-NIMH: Genetic Association Information Network – National institute of Mental Health

WTCCC: Wellcome Trust Case Control Consortium

STEP-BD: Systematic Treatment Enhancement Program for Bipolar Disorder

TGEN: Translational Genomics Research Institute

ConLiGen: The International Consortium on Lithium Genetics

BoMA: Bonn-Mannheim

dbGaP: Database of Genotypes and Phenotypes

NCP: Non-Centrality Parameter

GRR: Genotype-Relative Risk

Figure 1

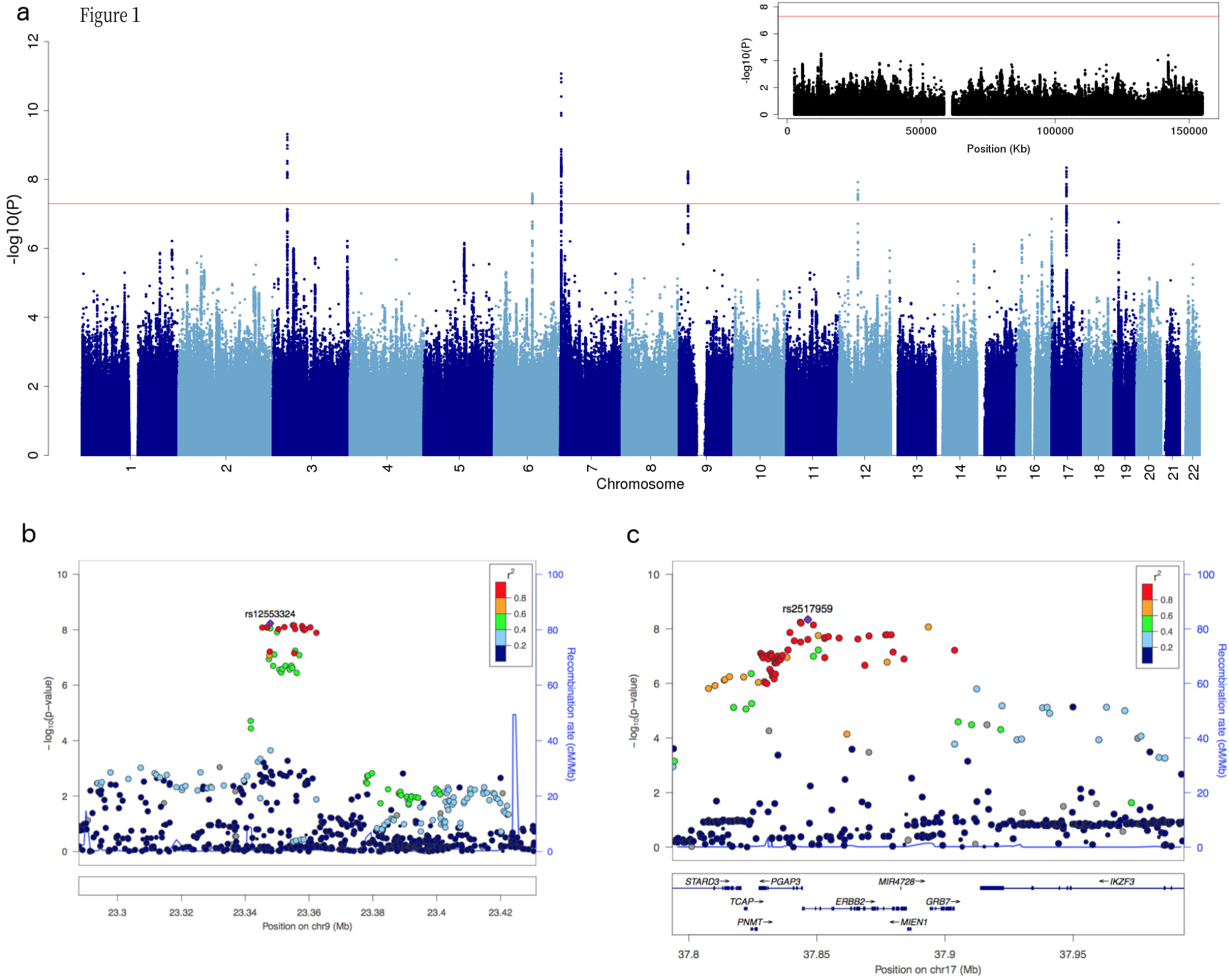
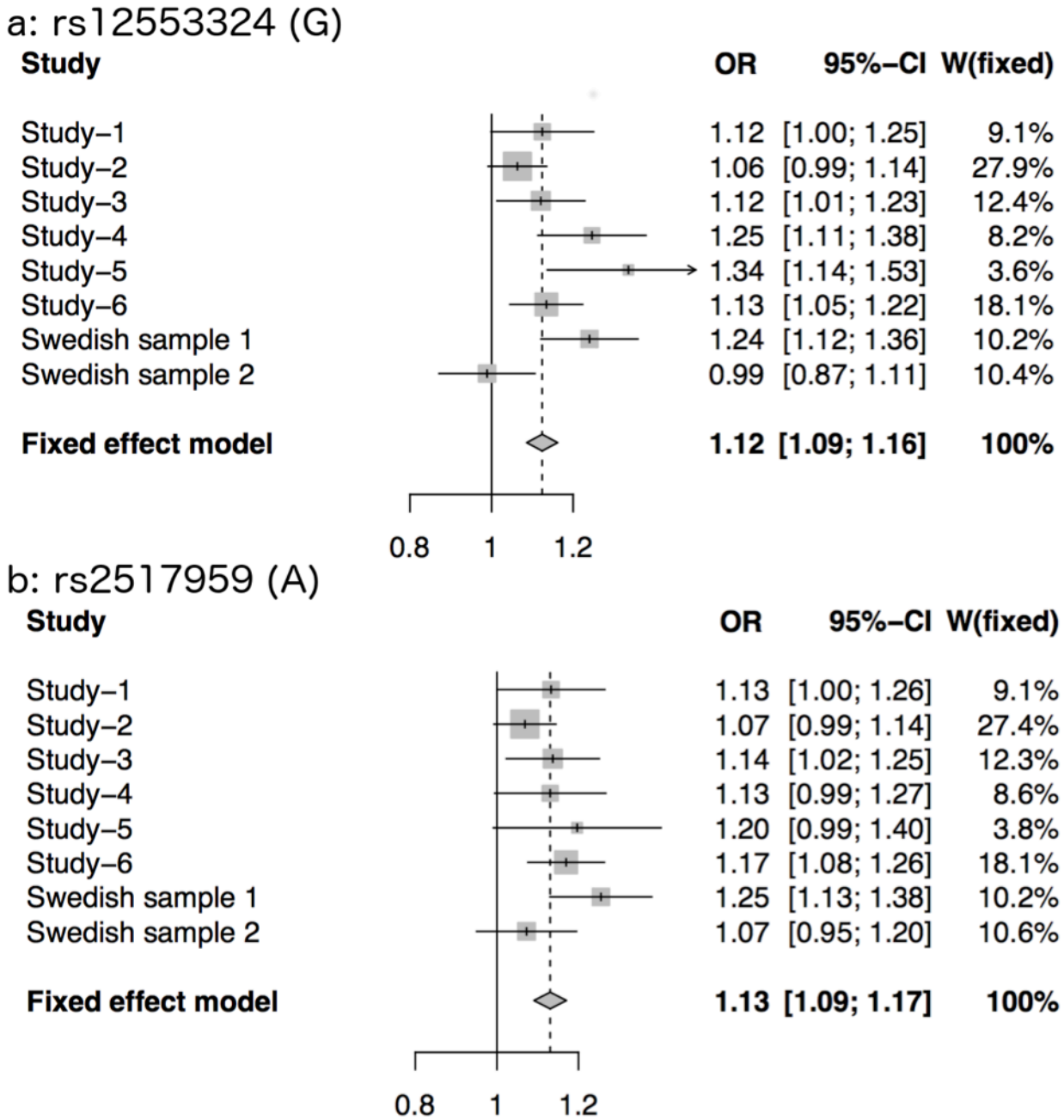

a
-log10(P)
Chromosome
Position (Kb)
b
rs12553324
Position on chr9 (Mb)
Recombination rate (cM/Mb)
c
rs2517959
STARD3
PGAP3
MIR4728
IKZF3
TCAP
ERBB2
GRB7
PNMT
MIEN1
Position on chr17 (Mb)

Figure 2


a: rs12553324 (G)
Study
OR
95%−CI
W(fixed)
Study−1 1.12 [1.00; 1.25] 9.1%
Study−2 1.06 [0.99; 1.14] 27.9%
Study−3 1.12 [1.01; 1.23] 12.4%
Study−4 1.25 [1.11; 1.38] 8.2%
Study−5 1.34 [1.14; 1.53] 3.6%
Study−6 1.13 [1.05; 1.22] 18.1%
Swedish sample 1 1.24 [1.12; 1.36] 10.2%
Swedish sample 2 0.99 [0.87; 1.11] 10.4%
Fixed effect model 1.12 [1.09; 1.16] 100%
0.8
1
1.2
b: rs2517959 (A)
Study
OR
95%−CI
W(fixed)
Study−1 1.13 [1.00; 1.26] 9.1%
Study−2 1.07 [0.99; 1.14] 27.4%
Study−3 1.14 [1.02; 1.25] 12.3%
Study−4 1.13 [0.99; 1.27] 8.6%
Study−5 1.20 [0.99; 1.40] 3.8%
Study−6 1.17 [1.08; 1.26] 18.1%
Swedish sample 1 1.25 [1.13; 1.38] 10.2%
Swedish sample 2 1.07 [0.95; 1.20] 10.6%
Fixed effect model 1.13 [1.09; 1.17] 100%
0.8
1
1.2