Published in final edited form as:

Mol Psychiatry. 2009 April ; 14(4): 359–375. doi:10.1038/mp.2008.125.

Genomewide Association for Major Depressive Disorder: A possible role for the presynaptic protein Piccolo

Patrick F. Sullivan, MD,
University of North Carolina, Chapel Hill pfsulliv@med.unc.edu

Eco J.C. de Geus, PhD,
VU University Amsterdam eco@psy.vu.nl

Gonneke Willemsen, PhD,
VU University Amsterdam ahm.willemsen@psy.vu.nl

Michael R. James, PhD,
Queensland Institute for Medical Research michael.james@qimr.edu.au

Jan H. Smit, PhD,
VU University Medical Center Amsterdam jh.smit@ggzba.nl

Tim Zandbelt, MSc,
VU University Medical Center Amsterdam timz@ggzba.nl

Volker Arolt, MD,

Correspond with Dr. Sullivan, Dept. of Genetics, CB#7264, 4109D Neurosciences Research Building, Univ. of North Carolina, Chapel Hill, NC, 27599-7264, USA. Voice: +919-966-3358, FAX: +919-966-3630. E-mail: pfsulliv@med.unc.edu..
†These authors contributed equally.

Financial Disclosures (Prior 3 Years)

Dr. Baune has received honoraria for educational training of psychiatrists and general practitioners from Lundbeck, Astra Zeneca and Pfizer Pharmaceuticals and travel grants from Astra Zeneca, Bristol-Meyer Squibb, Janssen and Pfizer Pharmaceuticals. Dr. Fava has received: research support from Abbott Laboratories, Alkermes, Aspect Medical Systems, Astra-Zeneca, Bristol-Myers Squibb Company, Cephalon, Eli Lilly & Company, Forest Pharmaceuticals Inc., GlaxoSmithkline, J & J Pharmaceuticals, Lichtwer Pharma GmbH, Lorex Pharmaceuticals, Novartis, Organon Inc., PamLab, LLC, Pfizer Inc, Pharmavite, Roche, Sanofi-Aventis, Solvay Pharmaceuticals, Inc., Synthelabo, Wyeth-Ayerst Laboratories; advisory/consulting fees from Abbott Laboratories, Amarin, Aspect Medical Systems, Astra-Zeneca, Auspex Pharmaceuticals, Bayer AG, Best Practice Project Management, Inc., Biovail Pharmaceuticals, Inc., BrainCells, Inc. Bristol-Myers Squibb Company, Cephalon, CNS Response, Compellis, Cypress Pharmaceuticals, Dov Pharmaceuticals, Eli Lilly & Company, EPIX Pharmaceuticals, Fabre-Kramer Pharmaceuticals, Inc., Forest Pharmaceuticals Inc., GlaxoSmithkline, Grunenthal GmBH, Janssen Pharmaceutica, Jazz Pharmaceuticals, J & J Pharmaceuticals, Knoll Pharmaceutical Company, Lorex Pharmaceuticals, Lundbeck, MedAvante, Inc., Merck, Neuronetics, Novartis, Nutrition 21, Organon Inc., PamLab, LLC, Pfizer Inc, PharmaStar, Pharmavite, Precision Human Biolaboratory, Roche, Sanofi-Aventis, Sepracor, Solvay Pharmaceuticals, Inc., Somaxon, Somerset Pharmaceuticals, Synthelabo, Takeda, Tetragenex, Transcept Pharmaceuticals, Vanda Pharmaceuticals Inc, Wyeth-Ayerst Laboratories; speaking fees from Astra-Zeneca, Boehringer-Ingelheim, Bristol-Myers Squibb Company, Cephalon, Eli Lilly & Company, Forest Pharmaceuticals Inc., GlaxoSmithkline, Novartis, Organon Inc., Pfizer Inc, PharmaStar, Primedia, Reed-Elsevier, Wyeth-Ayerst Laboratories; has equity holdings inCompellis, MedAvante; and has royalty/ patent, other income for patent applications for SPCD and for a combination of azapirones and bupropion in MDD, copyright royalties for the MGH CPFQ, DESS, and SAFER. Dr. Nolen has received: speaking fees from Astra Zeneca, Eli Lilly, Pfizer, Servier, Wyeth; unrestricted research funding from Astra Zeneca, Eli Lilly, GlaxoSmithKline, Wyeth; and served on advisory boards for Astra Zeneca, Cyberonics, Eli Lilly, GlaxoSmithKline, Pfizer, Servier. Dr. Perlis has received consulting fees or honoraria from AstraZeneca, Bristol Myers-Squibb, Eli Lilly, GlaxoSmithKline, Pfizer, and Proteus; he is a stockholder in Concordant Rater Systems, LLC, and the holder of a patent related to the monitoring of raters in clinical trials. Dr. Smoller has consulted to Eli Lilly, received honoraria from Hoffman-La Roche, Inc, Enterprise Analysis Corp. and MPM Capital, and has served on an advisory board for Roche Diagnostics Corporation. Dr. Sullivan has received unrestricted research support from Eli Lilly.

University of Münster volker.arolt@ukmuenster.de

Bernhard T. Baune, MD MPH,
James Cook University Queensland bernhard.baune@jcu.edu.au

Douglas Blackwood, MD,
University of Edinburgh d.blackwood@ed.ac.uk

Sven Cichon, PhD,
University of Bonn sven.cichon@uni-bonn.de

William L. Coventry,
University of New England wcovntr@une.edu.au

Katharina Domschke, MD MA,
University of Münster katharina.domschke@ukmuenster.de

Anne Farmer, MD PhD,
Institute of Psychiatry a.farmer@iop.kcl.ac.uk

Maurizio Fava, MD,
Harvard Medical School mfava@partners.org

Scott D. Gordon,
Queensland Institute for Medical Research scott.gordon@qimr.edu.au

Qianchuan He,
University of North Carolina, Chapel Hill heqianch@email.unc.edu

Andrew Heath, PhD,
Washington University, St. Louis heatha@psychiatry.wustl.edu

Peter Heutink, PhD,
VU University Medical Center Amsterdam p.heutink@vumc.nl

Florian Holsboer, MD PhD,
Max-Planck Institute of Psychiatry holsboer@mpipsykl.mpg.de

Witte J. Hoogendijk, MD PhD,
VU University Medical Center Amsterdam witteh@ggzba.nl

Jouke Jan Hottenga, PhD,
VU University Amsterdam jj.hottenga@psy.vu.nl

Yijuan Hu,
University of North Carolina, Chapel Hill yhu@bios.unc.edu

Martin Kohli, MSc,
Max-Planck Institute of Psychiatry kohlim@mpipsykl.mpg.de

Danyu Lin, PhD,
University of North Carolina, Chapel Hill lin@bios.unc.edu

Suzanne Lucae, MD PhD,
Max-Planck Institute of Psychiatry lucae@mpipsykl.mpg.de

Donald J. MacIntyre, MRCPsych,
Royal Edinburgh Hospital donald.macintyre@nhslothian.scot.nhs.uk

Wolfgang Maier, MD,
University of Bonn wolfgang.maier@ukb.uni-bonn.de

Kevin A. McGhee, PhD,
University of Edinburgh kevin.mcghee@ed.ac.uk

Peter McGuffin, MD PhD,
Institute of Psychiatry p.mcguffin@iop.kcl.ac.uk

Grant Montgomery, PhD,
Queensland Institute for Medical Research grant.montgomery@qimr.edu.au

Walter J. Muir, FRCPsych,
University of Edinburgh walter.muir@ed.ac.uk

Willem Nolen, MD,
University Medical Center Groningen w.a.nolen@psy.umcg.nl

Markus M. Nöthen, MD,
University of Bonn markus.noethen@uni-bonn.de

Roy H. Perlis, MD MSc,
Harvard Medical School rperlis@chgr.mgh.harvard.edu

Katrina Pirlo,
Institute of Psychiatry katrina.pirlo@iop.kcl.ac.uk

Danielle Posthuma, PhD,
VU University Amsterdam danielle@psy.vu.nl

Marcella Rietschel, MD PhD,
University of Heidelberg marcella.rietschel@zi-mannheim.de

Patizia Rizzu, PhD,
VU University Medical Center Amsterdam p.rizzu@vumc.nl

Alexandra Schosser, MD,
Institute of Psychiatry alexandra.schosser@iop.kcl.ac.uk

August B. Smit, PhD,
VU University Amsterdam guus.smit@falw.vu.nl

Jordan W. Smoller, MD ScD,
Harvard Medical School jordan_smoller@hms.harvard.edu

Jung-Ying Tzeng, PhD,
North Carolina State University jytzeng@stat.ncsu.edu

Richard van Dyck, MD PhD,
VU University Medical Center Amsterdam r.van.dyck@ggzba.nl

Matthijs Verhage, PhD,

VU University Amsterdam matthijs@cncr.vu.nl

Frans G. Zitman, MD PhD,
Leiden University Medical Center f.g.zitman@lumc.nl

Nicholas G. Martin, PhD,
Queensland Institute for Medical Research nick.martin@qimr.edu.au

Naomi R. Wray, PhD,
Queensland Institute for Medical Research naomi.wray@qimr.edu.au

Dorret I. Boomsma, PhD[†], and
VU University Amsterdam dorret@psy.vu.nl

Brenda W.J.H. Penninx, PhD[†]
VU University Medical Center Amsterdam b.penninx@vumc.nl

Abstract

Major depressive disorder (MDD) is a common complex trait with enormous public health significance. As part of the Genetic Association Information Network (GAIN) initiative of the US Foundation for the National Institutes of Health, we conducted a genomewide association study of 435,291 SNPs genotyped in 1,738 MDD cases and 1,802 controls selected to be at low liability for MDD. Eleven of the top 200 signals localized to a 167 kb region overlapping the gene piccolo (*PCLO*, whose protein product localizes to the cytomatrix of the presynaptic active zone and plays an important role in monoaminergic neurotransmission in the brain) with p-values of 7.7×10^{-7} for rs2715148 and 1.2×10^{-6} for rs2522833. We undertook replication of SNPs in this region in 5 independent samples (6,079 MDD independent cases and 5,893 controls) but no SNP exceeded the replication significance threshold when all replication samples were analyzed together. However, there was heterogeneity in the replication samples, and secondary analysis of the original sample with the sample of greatest similarity yielded $p=6.4\times10^{-8}$ for the non-synonymous SNP rs2522833 that gives rise to a serine to alanine substitution near a C2 calcium-binding-domain of the PCLO protein. With the integrated replication effort, we present a specific hypothesis for further studies.

Keywords

major depressive disorder; genome-wide association; Netherlands Study of Depression and Anxiety; Netherlands Twin Registry

Introduction

The defining features of major depressive disorder (MDD) are marked and persistent dysphoria plus additional cognitive signs and symptoms (anhedonia, sleep disturbance, weight/appetite changes, motor agitation/retardation, anergia, excessive guilt or worthlessness, poor concentration or indecisiveness, and recurrent thoughts of death or suicide) (1). MDD is distinct from normal sadness by its persistence (i.e., ≥2 weeks), additional signs and symptoms, and substantial associated impairment. The definition of MDD excludes other conditions typified by substantial depressive symptoms (other

psychiatric disorders, drug/alcohol dependence, and somatic diseases). The lifetime prevalence of MDD is ∼15% (2-4) and is twofold higher in women (5) with a course typified by recurrence of illness (6). It is associated with considerable morbidity (7-9), excess mortality from suicide and other causes (10-13), and substantial direct and indirect costs (14). A WHO study projected MDD to be the second leading cause of disability worldwide by 2020 (15).

Although there is a considerable corpus of research on the epidemiology and biological correlates of MDD, little is known for certain about its etiology. An important etiological clue may be the familial tendency of MDD and its heritability of 31-42% (16). This clue led to a number of genomewide linkage studies (Supplemental Methods) and studies of >100 theoretical or positional candidate genes. As for the use of these study designs with other biomedical disorders, their application to MDD has not been as successful as had been hoped.

It is now clear that genomewide association studies (GWAS) can be a successful tool in the genetic dissection of complex biomedical disorders (17, 18). The goal of this report is to describe a GWAS for MDD that was systematically designed to remediate a set of methodological issues common to genetic studies of MDD. Examples of these issues include small sample sizes, inhomogeneous samples in terms of ancestry and phenotyping, convenience sampling, and controls that are unaffected but not at low liability for MDD. Moreover, large-scale replication was integral to our design.

Methods

This GWAS was one of the six initial Genetic Association Information Network (GAIN) studies sponsored by the Foundation for the NIH (19). Individual phenotype and genotype data are available to researchers via application to the dbGaP repository (20). We have attempted to follow published guidelines for GWAS (21, Box 1).

Subjects

The parent projects that supplied subjects for this GWAS are longitudinal studies, the Netherlands Study of Depression and Anxiety (NESDA, http://www.nesda.nl) (22) and the Netherlands Twin Registry (NTR, http://www.tweelingenregister.org) (23). Sampling and data collection characteristics of the GAIN-MDD study have been described in detail elsewhere (24).

MDD cases were mainly from NESDA, a longitudinal cohort study designed to be representative of individuals with depressive and/or anxiety disorders. Recruitment of participants for NESDA took place from 09/2004-02/2007, and ascertainment was from outpatient specialist mental health facilities and via primary care screening. Additional cases were from the population-based cohorts NEMESIS (25), ARIADNE (26), and the NTR. Regardless of recruitment setting, similar inclusion and exclusion criteria were used to select MDD cases. Inclusion criteria were a lifetime diagnosis of DSM-IV MDD (1) as diagnosed via the Composite International Diagnostic Interview psychiatric interview (27), age 18-65 years, and self-reported western European ancestry. Persons who were not fluent in Dutch

and those with a primary diagnosis of schizophrenia or schizoaffective disorder, obsessive compulsive disorder, bipolar disorder, or severe substance use dependence were excluded (the etiology of MDD in these subjects may be distinct). The 1,862 cases included in GAIN were recruited from mental health care organizations (N=785), primary care (N=603), and community samples (NEMESIS N=218, ARIADNE N=96 and NTR N=160).

Control subjects were mainly from the NTR, which has collected longitudinal data from twins and their families since 1991 (total cohort of ∼22,000 participants from 5,546 families). The majority of families were recruited when the twins were adolescents or young adults through city council registrations along with alternative efforts to recruit older twins. Longitudinal phenotyping includes assessment of depressive symptoms (via multiple instruments), anxiety, neuroticism, and other personality measures. Inclusion required availability of both survey data and biological samples, no report of MDD at any measurement occasion, and low genetic liability for MDD. No report of MDD was determined by specific queries about medication use or whether the subject had ever sought treatment for depression symptoms and/or via the CIDI interview. Low genetic liability for MDD was determined by the use of a factor score derived from longitudinal measures of neuroticism, anxiety, and depressive symptoms (28) (mean 0, std 0.7); controls were required never to have scored highly (≥0.65) on this factor score. Finally, controls and their parents were required to have been born in the Netherlands or western Europe. Only one control per family was selected. There were N=1,703 controls from the NTR and additional controls from NESDA (N=133 from general practice, N=24 from ARIADNE). NESDA controls had no lifetime diagnosis of MDD or an anxiety disorder as assessed by the CIDI and reported low depressive symptoms at baseline (K-10 score <16 and Inventory of Depressive Symptoms score <4) (29, 30).

Case-Control Matching

If there were multiple eligible NTR controls in a family, we first matched on sex and age, and used the highest number of completed questionnaires as an additional criterion. Again, only one control per family was included.

DNA Sampling

Before the start of the NESDA and NTR biological sample collection, processing, and storage protocols were harmonized and DNA extraction was conducted concurrently in the same laboratory. For NESDA, blood sampling for the NESDA participants took place during the baseline visit (between 0830-0930) and DNA was isolated using the FlexiGene DNA AGF3000 kit (Qiagen, Valencia, CA, USA) on an AutoGenFlex 3000 workstation (Autogen, Holliston, MA, USA). For NTR, biological samples were taken in the subject's home (between 0700-1000) and DNA was extracted using the Puregene DNA isolation kit (Gentra, Minneapolis, MN, USA) for frozen whole blood samples. DNA concentrations were determined using the PicoGreen dsDNA Quantitation kit (Invitrogen Corporation, Carlsbad, CA, USA). All procedures were performed according to the manufacturer's protocols.

Ethical Issues

The NESDA and NTR studies were approved by the Central Ethics Committee on Research Involving Human Subjects of the VU University Medical Center, Amsterdam, an Institutional Review Board certified by the US Office of Human Research Protections (IRB number IRB-2991 under Federal wide Assurance-3703; IRB/institute codes, NESDA 03-183; NTR 03-180). All subjects provided written informed consent. As part of the GAIN application process, consent forms were specifically re-reviewed for suitability for the deposit of de-identified phenotype and genotype data into the controlled-access dbGaP repository (20). NESDA and NTR subjects were informed of participation in GAIN via newsletters. Only 22 NESDA respondents refused informed consent for genetic research (1.7% of all respondents approached).

GWAS Genotyping

Individual genotyping was conducted by Perlegen Sciences (Mountain View, CA, USA) using a set of four proprietary, high-density oligonucleotide arrays. The SNPs on these arrays were selected to tag common variation in the HapMap European and Asian panels using previously described genotype data (31), tagging approach (32), and methodology (33). At the beginning of GAIN, all HapMap (34) samples were genotyped with the Perlegen GWAS platform. Independent review of these data by the GAIN analysis group (19) showed 99.8% agreement with prior HapMap genotypes and the mean maximum r^2 between the Perlegen SNPs and HapMap phase II SNPs (31) was 0.89 for single and 0.96 for multi-marker analyses. The genotyping procedures and genotyping calling algorithms are described in the Supplemental Methods and in prior reports (35, 36). Briefly, 40 × 96-well plates were sent to Perlegen for GWAS genotyping. Genotyping was conducted blind to case-control status. Cases and controls were randomly allocated to plates and to positions within plates. Each plate contained DNA samples from 93 Dutch subjects plus 3 quality control samples. The three quality control samples included: two parents of one control on that plate (40 complete trios in total); and half the plates contained the same HapMap CEU sample (used for quality control in all GAIN projects) and half had a randomly-selected duplicate case sample. The total number of samples was 3,840 (=40 plates × 96 samples per plate) or 1,860 cases + 1,860 controls + 80 parents + 20 duplicate samples + 20 HapMap samples.

Quality Control – Subjects

Of the 3,820 Dutch samples sent to Perlegen (excluding the 20 HapMap internal control samples), genotypes were delivered for 3,761 samples. Fifty-nine samples did not have GWAS data: 39 samples with uncertain linkage between genotype and phenotype records, 7 samples with evidence of contamination, 6 samples that failed genotyping, and 7 miscellaneous failures (2 of these were excluded as chrX and chrY genotyping data were consistent with the presence of XO and XXY sex chromosome status). After further analysis, 8 subjects were removed for excessive missing genotype data (>25%), 1 case for high genomewide homozygosity (∼75%), 38 subjects whose genomewide IBS estimates were consistent with first- or second-degree relationships, and 57 additional subjects whose ancestry diverged from the remainder of the sample (see Supplemental Methods for details).

After these exclusions (N=104) and removing duplicated and trio quality control sample, there were 3,540 subjects in the final analysis dataset including 1,738 cases and 1,802 controls. The principal reason for fewer cases than controls was the higher prevalence of substantial non-European ancestry. The list of subjects in the final analyses dataset is included as a Supplemental File ("mddC.fam").

Quality Control – SNPs

The unfiltered dataset obtained from dbGaP contained 599,156 unique SNPs. The Perlegen genotyping algorithm yielded a quality score for each individual genotype, and a more stringent quality score cutoff (≥ 10) than that used by Perlegen was applied. The SNP quality control process is described in detail in the Supplemental Methods. Briefly, to be included in the final analysis dataset, SNPs were required not to have any of the following features: gross mapping problem (37), ≥2 genotype disagreements in 40 duplicated samples, ≥2 Mendelian inheritance errors in 38 complete trio samples, minor allele frequency <0.01, or >0.05 missing genotypes in either cases or controls. A Hardy-Weinberg filter was not used as lack of fit to Hardy-Weinberg expectations can occur for valid reasons (e.g., a true association) (38) and given that 95.6% (=51,592/53,994) of SNPs with $p<0.00001$ from an exact test of Hardy-Weinberg equilibrium (39) in controls were already flagged for exclusion. A total of 435,291 SNPs met these criteria and were included in the final analysis dataset (included as a Supplemental File, "mddC.bim"). Additional quality control checks are described in the Supplemental Methods). Thirteen controls were genotyped in a different study using the Illumina 317K platform and, of the 82,636 SNPs common to both platforms, the genotype agreement was 99.94%.

Single Marker Statistical Analyses

There were three classes of SNPs – those that could be heterozygous in all subjects (chr1-22 and chrX/PAR1), those that were heterozygous in females (non-PAR chrX), and those that were hemizygous in males (non-PAR chrX and chrY). All SNPs that passed quality control checks were tested for association with MDD using 1 df Cochran-Armitage trend tests. For complex traits, it is widely believed that the contributions of individual SNPs to disease risk are often roughly additive (40) . The Cochran-Armitage trend test can be used to detect such effects. This test is usually recommended due to its robustness to the violation of the HWE assumption (41): P-values from females and males for non-PAR chrX were combined using Fisher's method (42).

Population stratification artifacts were assessed in two ways. As described elsewhere (36), including principal components as covariates in a logistic regression model can robustly control stratification effects. To do this, we identified a set of 127,688 SNPs in linkage equilibrium (43) and used the "smartpca" program in EigenSoft (44) to compute 10 principal components for each subject that were included as covariates in logistic regression models (case/control status $\sim SNP_i + PC1 + PC2 + \ldots + PC10$). We also used a stratified Cochran-Mantel-Haenszel test in PLINK (43) as a complementary approach.

For noteworthy associations, there were additional checks to ensure that an association was not due to experimental bias. These checks included: manual inspection of SNP cluster plots

to ensure reasonable performance of the genotyping calling algorithm; evaluation of conformation to Hardy-Weinberg equilibrium in controls, cases, and overall (discussed in the Supplemental Methods); the checks for population stratification described above; evaluation of plate-specific association results to ensure that the overall association was not driven by one or a few plates; comparison of control MAFs to the HapMap EUR panel; and evaluation of the characteristics of a SNP in high linkage disequilibrium ("proxy association") as a similar association with such a SNP decreases the chance of some forms of method artifacts.

Control of False Discoveries

Given the 10^5-10^7 statistical comparisons in a GWAS, small p-values are expected by chance. To control the risk of false discoveries, q-values (45, 46) were computed for all p-values for single-marker tests of association. A q-value is an estimate of the proportion of false discoveries among all significant markers, or the false discovery rate (FDR) for the corresponding p-value. The use of q-values is preferable to more traditional multiple testing controls because q-values provide a better balance between the competing goals of finding true positives versus controlling false discoveries, allow more similar comparisons across studies because proportions of false discoveries are much less dependent on the number of tests conducted and are relatively robust against the effects of correlated tests (45, 47-54). The q-value threshold for declaring significance was 0.10 (i.e., the top 10% of the significant findings are, on average, allowed to be false discoveries) (50, 55). FDR thresholds <0.10 result in a disproportionate drop in power to detect true effects.

Imputation

We used two imputation approaches, the SNPMStat method of Lin et al. (56) to impute 246 additional SNPs in the *PCLO* region and Abecasis' MACH (v1) to impute 2,037,829 autosomal SNPs with $R^2 \geq 0.5$ (a cutoff that removes ~90% of SNPs with unreliable imputation results while sacrificing 2-3% of reliably imputed SNPs). Both SNPMStat and MACH gave similar results in the *PCLO* region. Imputed genotypes were used in secondary analyses. The HapMap2 EUR panel (31, 34) was used as reference.

Statistical Power

Quanto (57, 58) was used to approximate statistical power given the following assumptions: two-tailed $\alpha=1\times10^{-7}$ (=0.05/500,000), 1,738 cases and 1,802 controls, lifetime morbid risk of MDD of 0.15, and a log additive genetic model. For statistical power of 0.80 ($\beta=0.20$), the minimum detectable genotypic relative risks are 1.59, 1.40, and 1.35 for minor allele frequencies of 0.10, 0.25, and 0.40.

Software

PLINK (v1.0) (43), SAS (v9.1.3) (59), R (v2.6.1) (60), HAPSTAT (v3) (61-63), MACH1, SNPMStat (56), HaploView (64), and JMP (v6) (65) were used for data management, quality control, statistical analyses, and graphics.

Bioinformatics

All genomic locations are per NCBI Build 35 (66) (UCSC hg17) (67). Pseudo-autosomal region 1 (PAR1) is assumed to be located on chrX:1-2,692,881 and chrY:1-2,692,881 and PAR2 on chrX:154,494,747-154,824,264 and chrY:57,372,174-57,701,691 (68). SNP annotations were per TAMAL (37) based chiefly on UCSC genome browser files (67), HapMap (34), and dbSNP (66).

Results

Sample Description

Table 1 presents descriptive data for cases and controls. Controls had a higher proportion of males and were slightly older (and thus were farther through the period of risk for MDD). Consistent with known correlates of MDD, cases had a significantly lower educational level, less often had a partner, were more often smokers, and scored much higher on the NEO-FFI neuroticism scale.

SNP Description

The analysis SNP set had 435,291 SNPs including 427,049 autosomal SNPs, 7.988 SNPs on the non-PAR portions of chrX, 239 SNPs on chrXY/PAR1, 15 SNPs on chrY, and 0 SNPs on PAR2. The median SNP missingness was 0.00339 (inter-quartile range 0.00113-0.0105) and the median minor allele frequency was 0.2422 (inter-quartile range 0.1375-0.3646) with similar estimates in cases and controls. The average marker density over the genome was 1 SNP every 7,069 bases (=3,077,088,087 bases / 435,291 SNPs). The median inter-marker distance was 2,911 bases with interquartile range 966-7,374 bases and a 99th percentile of 50.1 kb.

Single Marker Association Tests

We used the Cochran-Armitage trend test to test for association of the 435,291 SNPs in the GWAS dataset with case/control status. The estimated lambda (69) was 1.046 (similar p-value minima and lambdas were obtained using logistic regression with 10 principal components and using a stratified Cochran-Mantel-Haenszel tests based on identity-by-state clusters) (43, 44). The minimum q-value was 0.28 (i.e., if these tests were called significant, over the long-term, a minimum false discovery rate of ∼28% would be incurred). As the pre-specified q-value threshold was 0.10, no SNP reached genomewide significance. The proportion of all SNPs without true effects (p_0) (54) was conservatively estimated to be p_0=0.9999954, consistent with the presence of ∼2 SNPs with true effects in these GWAS data. The results of this MDD GWAS are included as a Supplemental File (“results.txt”) to facilitate comparisons with other studies.

Panel a of Figure 1 depicts the quantile-quantile plots (40) for these analyses. The observed p-values do not strongly depart from the p-value distribution expected by chance. Panel b of Figure 1 shows a plot of $-\log_{10}(p_{trend})$ by genomic location.

Table 2 presents the findings for the top 25 SNPs. The quality control metrics – SNP missingness, agreement with HWE, and similarity of the control MAFs to the HapMap EUR

panel – for the top 25 SNPs are generally acceptable. Four of the top 25 associations are in the presynaptic cytomatrix protein piccolo (*PCLO*). Table 3 depicts the top 25 multi-SNP clusters (i.e., for an index SNP with association p<0.001, these clusters are additional SNPs within 250 kb of the index SNP with with $r^2 \geq 0.50$). The full version of this table is included as a Supplemental File ("Table3_full.xls"). *PCLO* is present in the top 25 clusters along with two additional multi-SNP clusters in the top 200. Other notable SNP clusters occurred in *GRM7* (rank 51), *DGKH* (rank 83, a candidate gene for bipolar disorder) (70), *DAOA* (rank 124), and *DRD2* (rank 226).

Focusing on Piccolo (*PCLO*)

Although no association met genomewide significance, there were clusters of SNPs in *PCLO* (Figure 2). Notably, 11 of the 200 smallest p-values localized to a 167 kb segment overlapping *PCLO*. Interest in *PCLO* was increased given its expression in brain, localization to the presynaptic active zone (71), and involvement in monoamine neurotransmission, a venerable hypothesis of the etiology of MDD (72). Moreover, the third most significant SNP (rs2522833) codes for a non-synonymous amino acid change (ala-4814-ser) in *PCLO* near its C2A calcium binding domain (73).

We investigated possible causes of spurious associations in the *PCLO* region (chr7:82,032,093-82,436,848). First, these findings were not due to plate effects as inspection of plate-specific association data for these SNPs did not show any marked outliers or systematic biases. Second, review of allelic intensity cluster plots on which genotype calls were based revealed adequate performance of the Perlegen genotype calling algorithm. Third, inspection of additional quality control metrics did not suggest systematic problems with SNPs in this region. Fourth, inspection of LD matrices excluded very high LD as the sole explanation for the results (Supplemental Figure 10), and none of the genotyped SNPs had strong LD ($r^2 \geq 0.8$) with rs2715148 (the SNP with the smallest p-value in the *PCLO* region). Fifth, population stratification can cause false positive findings but this did not appear to explain the *PCLO* association: (a) the same 11 SNPs had p-values among the top 200 associations in unadjusted analyses as well as with adjustment via principal components and stratified analyses; and (b) for the 57 SNPs in the *PCLO* region, the p-values across these three types of analyses were consistent (the Spearman correlations between p-values from trend tests, logistic regression, and stratified analyses were all >0.962). Sixth, the minor allele frequencies in the control group in the *PCLO* region were usually quite similar to available EUR control groups suggesting that the *PCLO* findings were not due to an artifact of the control selection process (see below). Finally, bioinformatic investigation did not suggest that this is a problematic region to genotype as the *PCLO* region is not known to be under positive selection in humans (74), to contain segmental duplications (67), or common copy number variants (search of the Database of Genomic Variants yielded two rare CNVs with control frequencies of 0.12% and 0.89%) (75, 76 , 77).

We conducted additional analyses to attempt to localize the association depicted in Figure 2. Imputation (56) supported the directly typed SNP associations but did not yield an association p-value markedly more significant than any directly genotyped SNP (although

22 of the 25 most significant imputed associations in the genome were in this region). Haplotype analysis using 3-SNP sliding windows did not improve localization. Secondary analyses by sex, case ascertainment setting, and recurrent early-onset MDD (reoMDD, arguably the most heritable form of MDD) (16, 78) suggested that most of the signal was from females and from subjects with reoMDD (Supplemental Table 11). The findings for reoMDD were often stronger than the primary analyses, particularly for the most significant SNP (rs2715148) where the p-value decreased by 1.2 orders of magnitude to 9.5×10^{-8}.

PCLO Replication

Although no finding met genomewide significance, the presence of multiple possible signals in *PCLO* and the plausibility of a role for *PCLO* in the etiology of MDD led us to attempt replication in external samples. We assembled a collection of 11,972 independent subjects (6,079 MDD cases and 5,893 controls) from seven different groups and a total of six case-control replication samples (two German samples were combined, see Supplemental Methods). As with NESDA cases, all replication cases were adults of European ancestry on whom a structured clinical interview was used to substantiate the lifetime diagnosis of DSM-IV MDD (1), and all studies excluded common MDD phenocopies (e.g., depressive symptoms due to another psychiatric disorder or a general medical condition). As with NTR controls, all replication controls were adults of European ancestry ascertained from the population, and individuals reporting MDD symptoms were excluded. We estimated statistical power using Quanto (57) (assumptions: log-additive genetic model, MDD lifetime risk 0.15, MAF=0.45 (similar to rs2522833), a genotypic relative risk of 1.14 (“shrunk” down from the observed GRR of 1.26 for rs2522833 to account for the “Winner's Curse” phenomenon) (79), and a conservative two-tailed type 1 error rate of 0.00167 (=0.05/30 replication SNPs). Statistical power was 97.2% for replication for the two SNPs genotyped in all samples (N=11,972) and 90.4% for the remaining SNPs (N=9,278). Five replication samples were genotyped for 30 SNPs using the same Sequenom iPlex SNP pool (15 SNPs were in the primary GWAS and 15 were selected to tag common variation in Europeans) (80) and one sample was successfully genotyped for two SNPs using TaqMan. The SNP selection strategy effectively cast a broad net over the region showing association in Figure 2. For the NESDA/NTR samples, agreement between the initial Perlegen genotypes in this region and independent re-genotyping was high (0.9987).

The single SNP results for MDD are depicted in Figure 3 and Table 4a. Our analytic plan dictated the combined analysis of all replication samples with the use of a one-tailed directional test. No association in the replication sample reached statistical significance after correction for multiple comparisons and SNP non-independence due to LD (ninth column in Table 4a). Similarly, haplotype analyses did not reveal significantly associated regions (Supplemental Figure 16). There were four p-values <0.05 in the replication sample but only rs10954694 also had Z-scores of the same sign in both samples. Table 4b shows the results for reoMDD, and no single SNP was significant after correction for multiple comparisons. When we repeated the MDD analyses restricted to female subjects, the observed significance levels did not become markedly stronger in any of the replication samples in contrast to the initial NESDA/NTR sample. Thus, results from analyses of all replication

samples did not reach the *a priori* criterion for replication evidence for the involvement of *PCLO* in the etiology of MDD.

Unanticipated heterogeneity in cases

However, we observed, *a posteriori*, that there was potentially important heterogeneity in the replication samples for eight SNPs that were strongly associated in the original sample ($I^2 \geq 0.4$, ninth column in Table 4a). In investigating this further (Supplemental Methods), we determined that there was little evidence for genetic heterogeneity in the genotyped region for controls but, unexpectedly, there was significant heterogeneity in the cases. Principal components analysis and inspection of Table 4a and the forest plots in Figure 3 indicated that the outlier was the Australian QIMR sample. Notably, the original and QIMR samples were particularly similar in that both studies included population-based cases and controls were selected to be at low liability for MDD based on longitudinal assessments. Of the nine SNPs with $p<0.05$ in the QIMR sample, eight had both low p-values and Z-scores with the same sign as in the NESDA/NTR sample. As an exploratory analysis, we analyzed the original and QIMR samples jointly, and the minimum p-value was 6.4×10^{-8} at the non-synonymous SNP rs2522833 that gives rise to a serine to alanine substitution near the C2A calcium-binding-domain of the PCLO protein.

Secondary analyses

We conducted additional analyses of the NESDA/NTR GWAS dataset that were specified *a priori* but which should be considered exploratory.

(a) The network of proteins with which *PCLO* interacts in its role at the presynaptic cytoskeletal matrix is relatively well-characterized, and we reasoned that genes encoding these proteins might harbor risk or protective variants. We assessed this hypothesis by testing for association conditioning on the *PCLO* nsSNP rs2522833 (i.e., investigating whether controlling statistically for the effect of rs2522833 increases the salience of other SNP associations), assessing the minimum p-value per gene, and then comparing this list to a list of 54 genes that make proteins that interact with *PCLO*. This analysis did not reveal any SNPs or genes whose significance was markedly lower than without including rs2522833 in the logistic regression model. Moreover, no known *PCLO* interacting protein was notable on this list.

(b) We imputed genotypes for 2,037,829 autosomal SNPs using MACH with reference to HapMap CEU genotypes. The resulting lambda was 1.048, and the minimum p-value was 1.21×10^{-7}. As noted above, 22 of the 25 most significant imputed associations were in the *PCLO* region. Investigation of SNP clustering that accounted for LD yielded results similar to those shown in Table 3.

(c) We assembled a list of 103 candidate genes that had been studied for association with MDD in the literature (81). Nineteen of these genes had no SNPs within its transcript and another 9 genes had inadequate coverage (>1 SNP/15 kb; Supplemental Table 17). Of the remaining 75 genes, only *NOS1* (neuronal nitric oxide synthase, $p=0.0006$) had $p<0.001$.

However, *NOS1* (as with most genes in Supplemental Table 16) is quite large and chance is a prominent potential influence on these results.

(d) We compared the GWAS association results to a meta-analysis of gene expression data from 12 studies of post-mortem brain tissue in MDD cases compared with controls (10 frontal cortex and 2 cerebellum studies). These data are available via the Stanley Foundation (http://www.stanleygenomic.org). There were five genes with GWAS p<0.05 (all had gene expression changes significant at p 0.0004 – 0.007). The genes were: *SGCG* (sarcoglycan), *CALD1* (caldesmon 1), *EEF1A1* (eukaryotic translation elongation factor 1 alpha 1), *CFLAR* (CASP8 and FADD-like apoptosis regulator), and *TP73L* (tumor protein p73-like). There is no overlap of this list with the *PCLO* interactors or MDD candidate genes from the literature.

(d) Alternative models, filters, and phenotypes. (i) For reoMDD, the minimum p-value over all GWAS SNPs was at the *PCLO* region SNP rs2715148 (8.4×10^{-8}) which ranked second of all SNPs using the trend test (Table 2). (ii) rs2715148 also had the smallest p-value under a dominant model of SNP action (6.2×10^{-6}). (iii) Given the female predominance in MDD, we analyzed data from females and males separately. For female cases and controls, rs2715148 had the smallest p-value (4.0×10^{-7}) and multiple other *PCLO* SNPs had p-values in the 10^{-5} – 10^{-6} range. For males, most *PCLO* SNPs had p > 0.05 and the minimum was in the *SLC9A9* SNP rs4839627 (9.1×10^{-7}). (iv) Again, given sex differences in MDD prevalence, we investigated SNPs on chrX and chrY more closely. The minimum p-value in chrX pseudo-autosomal region 1 was 0.02. For the non-PAR regions of chrX in females, the SNPs with the smallest p-values were rs11094388 (p=0.0003, intergenic), rs5971108 (p=0.0003, *PTCHD1*), rs5930667 (p=0.0004, intergenic), rs4618863 (p=0.0005, intergenic), rs2207796 (p=0.0005, in the very large gene *DMD*), and rs5936428 (p=0.0009, *FMR2*). For males, the minimum p-value on chrX was at rs10521594 ($p=5.4\times10^{-5}$, intergenic) and 0.22 on chrY.

Discussion

Overview

Major depressive disorder (MDD) is a common complex trait of enormous public health significance. As part of the GAIN initiative of the US Foundation for the NIH (19), we conducted a genomewide association study of 435,291 SNPs genotyped in 1,738 MDD cases and 1,802 controls selected to be at low liability for MDD. Our study had numerous positive attributes including its historically large sample size, its largely population-based and longitudinal design, and relatively unbiased and dense genomewide genotyping designed to capture common variation in subjects of European ancestry.

According to our primary analysis plan, no SNP-MDD phenotype association reached genomewide significance as the minimum q-value was 0.28, greater than the pre-defined q-value threshold of 0.10. This result was not unexpected. For example, type 2 diabetes mellitus has arguably reaped the greatest harvest from GWAS (82) and yet two of the initial T2DM GWAS were unremarkable when analyzed independently (83, 84). One of the key

lessons of the GWAS era is the importance of meta-analysis where its application to the primary GWAS can uncover positive findings that replicate well across studies (18, 85).

Is PCLO a causal risk factor for MDD?

Although no locus exceeded the genomewide threshold after correction for multiple comparisons, 11 of the top 200 signals localized to a 167 kb region overlapping the gene piccolo (*PCLO*). The protein product of *PCLO* localizes to the presynaptic active zone and plays an important role in brain monoaminergic neurotransmission (86), clearly intersecting with a venerable hypothesis of the etiology of mood disorders (87). Moreover, the third most significant association was a common non-synonymous SNP near its critical C2A binding domain in PCLO (88, 89). Although it is an obvious candidate gene, we are not aware of any prior association studies of *PCLO* and mood disorders (*PCLO* is in a region of 7q implicated by linkage in autism and one autism association study has been published) (90).

We judged the intersection of this GWAS result with prior knowledge sufficient to trigger a large-scale replication effort by genotyping *PCLO* SNPs in 6,079 MDD independent cases and 5,893 controls. Statistical power to replicate exceeded 90% even after accounting for (79) the “Winner's Curse” phenomenon (a form of regression to the mean whereby the true genotypic relative risk is over-estimated in the initial study) (91, 92). However, in spite of the apparent *a priori* strength of a hypothesis of genetic variation in *PCLO* in the etiology of MDD, no SNP analyzed in the replication sample met appropriately rigorous criteria for replication (21). Therefore, unlike GWAS for many non-psychiatric biomedical disorders, our GWAS and replication efforts did not yield “proof beyond a reasonable doubt” level of evidence for an association between genetic variation in *PCLO* and MDD.

Investigation of the sources of heterogeneity in the replication samples indicated that controls were genetically similar to the original sample in the *PCLO* region but that cases were dissimilar. We observed, *a posteriori*, that both principal components derived from *PCLO* region genotypes in QIMR cases and effect size estimates in the QIMR replication sample tended to be similar to the original sample. This is notable because, of all the replication samples, ascertainment of QIMR subjects was most similar to the primary NESDA/NTR sample in that cases were identified from population-based sources (100% for QIMR and 60% for NESDA) rather than tertiary sources as for the other replication samples. MDD cases from clinical samples may differ from population-based cases due to selection bias (93), Berkson's bias (94, 95), differing referral filters (96), or even a different genetic basis (97) with respect to genetic variation in the *PCLO* region.

Joint analysis of the NESDA/NTR and QIMR samples yielded $p=6.4x10^{-8}$ (uncorrected for multiple hypothesis testing) for the non-synonymous SNP rs2522833. This result suggests a specific hypothesis for future studies: an association between genetic variation in *PCLO* and MDD may be detected only in population-based cases. Thus, it would be premature to exclude *PCLO* from a role in the etiology of some forms of MDD.

The Heterogeneous Nature of MDD

Interpretation of the *PCLO* replication efforts is consistent with two broad possibilities. The first possibility is that genetic variation in *PCLO* is truly not associated with MDD. This interpretation is supported by the replication analyses (specified *a priori*) in which no SNP was significantly associated after correction for multiple comparisons and SNP dependence due to LD. This strict interpretation is generally viewed as “best practice” in human genetics (21) but implicitly assumes etiological homogeneity for MDD in the *PCLO* region. The second possibility invokes a less parsimonious model involving heterogeneity, that genetic variation in *PCLO* is etiologically causal to some subtypes of MDD. This interpretation is an *a posteriori* hypothesis consistent with the empirical results particularly in the notable differences in associations between samples, case ascertainment strategies, and indications from principal components analysis that NESDA and QIMR cases are more similar than the clinically ascertained subjects. It is notable that the control samples from each site were considerably more similar than cases from the same sites.

The tension between null *a priori* results and plausible *a posteriori* hypotheses is a core issue in psychiatric genetics. Important phenotypes like MDD are defined reliably and with reference to diagnostic schema developed principally for clinical purposes. Heterogeneous etiology of MDD is widely suspected but there are no proven ways to index heterogeneity (indeed, a prominent rationale for genetics studies is improve differential diagnosis).

Our results are consistent with prior observations of the heterogeneous nature of MDD, particularly with regard to ascertainment. Individuals who meet MDD criteria from community or primary care sources may have a more inclusive and less comorbid form of MDD (98) whereas tertiary ascertainment may yield subjects with greater comorbidity and perhaps distinctive etiology (99). In particular, it is formally possible (but unproven) that the *PCLO* results are accurate – genetic variation in *PCLO* might be causal to the types of MDD seen in community samples but other loci contribute to a distinctive type of MDD seen in tertiary care samples.

Other hypotheses

There were two MDD cases who may have had unrecognized genomic disorders (100) (possible Turner's and Klinefelter's syndromes). We speculate that small numbers of cases with MDD will have CNV-related genomic disorders that are plausibly causal to MDD. Clarification of the role of such rare variants will require larger samples.

Most of the additional exploratory analyses were unrevealing, including examination of proteins known to interact with PCLO, genotype imputation, comparison of GWAS findings with MDD candidate genes from the literature and gene expression changes in the brain in cases with MDD, and alternative genetic models, phenotype definitions, and sex-specific analyses.

We searched the SLEP compendium of psychiatric genetics findings (101) to attempt to discover overlap of our findings with those reported in the literature. First, with reference to a meta-analysis of microarray studies on the Stanley brain bank MDD and control samples, expression of *CFLAR* and *MARCH3* were increased and *LST1* and *HLA-B* were decreased in

MDD post-mortem frontal cortex. These regions ranked 9, 232, 267, and 432 in the NESDA/NTR GWAS. Second, we looked for convergence of our findings with other GWAS of psychiatric disorders. Notable genomic locations of overlap of the top 480 regions in the present GWAS were found with GWAS for ADHD (*ITIH1*, S Faraone, personal communication), the WTCCC GWAS for bipolar disorder (*SHFM1* and *UGT2B4*) (102), and a bipolar GWAS that used DNA pooling (*GRM7* and *DGKH*) (70). Third, we looked at the minimum p-values in our study for genes that met or nearly achieved genomewide significance: the minimum p-values in our study for *MAMDC1* (103) was 0.004, 0.03 for *ZNF804A* (104), 0.002 in *ANK3* (105), and 0.03 in *CACNA1C* (105). These overlaps are intriguing (although the play of chance cannot be excluded), and will be formally investigated as part of our participation in the Psychiatric GWAS Consortium analyses (18).

Limitations

(a) Although statistical power has been systematically underestimated in psychiatric genetics, when we began this study in Q3 2006, it was believed that statistical power would be reasonable to detect realistic genetic effects. However, the definition of “realistic” has shifted considerably since 2006 and it may be important to design studies that can detect genotypic relative risks <1.10. (b) When this study began, the coverage and performance of the Perlegen GWAS platform were among the better options available (19). The technology and pricing have evolved rapidly and superior platforms are now available. A key limitation of the Perlegen platform is its inability to assess copy number variation (106) that may be particularly salient for psychiatric disorders (107, 108). More generally, the GWAS platform might not be sufficiently “genomewide” and unbiased: the platform may have had inadequate coverage in an etiologically important region of the genome, SNPs are only one type of genetic variation, and important non-SNP genetic variation might not have been sufficiently well captured. (c) There was an imbalance in the proportion of males in cases and controls. Although it is unclear whether and how this might bias the results, it may have lead to some degree of bias. (d) Finally, GWAS are predicated upon the crucial assumption that the predominant diagnostic criteria are valid with respect to the fundamental architecture of the disorder.

Conclusions

We describe here a large effort to identify DNA sequence variation fundamental to MDD. Although our initial GWAS results for the *PCLO* region were intriguing, this highly plausible hypothesis did not find support in a large-scale replication attempt. Our hypothesis about a role of genetic variation in *PCLO* for MDD in population but not clinical settings emphasizes the importance of knowing the epidemiological sampling frame for a study. Finally, we hope that the model we used in this study – a cooperative international effort – will be adopted by groups studying other psychiatric disorders in order to maximize progress.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We acknowledge support from NWO: genetic basis of anxiety and depression (904-61-090); resolving cause and effect in the association between exercise and well-being (904-61-193); twin-family database for behavior genomics studies (480-04-004); twin research focusing on behavior (400-05-717), Center for Medical Systems Biology (NWO Genomics); Spinozapremie (SPI 56-464-14192); Centre for Neurogenomics and Cognitive Research (CNCR-VU); genomewide analyses of European twin and population cohorts (EU/QLRT-2001-01254); genome scan for neuroticism (NIMH R01 MH059160); Geestkracht program of ZonMW (10-000-1002); matching funds from universities and mental health care institutes involved in NESDA (GGZ Buitenamstel-Geestgronden, Rivierduinen, University Medical Center Groningen, GGZ Lentis, GGZ Friesland, GGZ Drenthe). Genotyping was funded by the Genetic Association Information Network (GAIN) of the Foundation for the US National Institutes of Health, and analysis was supported by grants from GAIN and the NIMH (MH081802). Genotype data were obtained from dbGaP (http://www.ncbi.nlm.nih.gov/dbgap, accession number phs000020.v1.p1). Statistical analyses were carried out on the Genetic Cluster Computer (http://www.geneticcluster.org) which is financially supported by the NWO (480-05-003). Dr. Sullivan was also supported by R01s MH074027 and MH077139. Dr. Schosser was supported by an Austrian Science Fund Erwin-Schrödinger-Fellowship. We would like to express our thanks to: the GAIN Genotyping group (Dr. Gonçalo Abecasis, chair) for help with quality control; Drs. Gonçalo Abecasis and Jun Li for assistance with MACH; Dr. Shaun Purcell for PLINK; Troy Dumenil (QIMR) for expert assistance with the replication genotyping; Dr. Dina Ruano (Portuguese Foundation for Science and Technology, SFRH/BPD/28725/2006); and Drs. Pam Madden (DA012854) and Richard Todd (AA013320) for supplying some of the phenotypes used in the Australian sample. Replication genotyping of the STAR*D samples was supported by a grant from the Bowman Family Foundation and the Sidney R. Baer, Jr. Foundation. We gratefully acknowledge NARSAD for funding the *PCLO* follow-up genotyping.

References

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. Fourth Edition ed.. American Psychiatric Association; Washington, DC: 1994.
2. Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshleman S, Wittchen H-U, Kendler KS. Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: results from the National Comorbidity Survey. Archives of General Psychiatry. 1994; 51:8–19. [PubMed: 8279933]
3. Kessler RC, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR, Rush AJ, Walters EE, Wang PS. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). Jama. 2003; 289(23):3095–105. [PubMed: 12813115]
4. Kessler RC, Ustun TB. The World Mental Health (WMH) Survey Initiative Version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). Int J Methods Psychiatr Res. 2004; 13(2):93–121. [PubMed: 15297906]
5. Weissman MM, Bland R, Joyce PR, Newman S, Wells JE. Wittchen H-U. Sex differences in rates of depression: cross-national perspectives. Journal of Affective Disorders. 1993; 29:77–84. [PubMed: 8300980]
6. Piccinelli M, Wilkinson G. Outcome of depression in psychiatric settings. British Journal of Psychiatry. 1994; 164:297–304. [PubMed: 8199782]
7. Wells KB, Stewart A, Hays RD, Burnam MA, Rogers W, Daniels M, Berry S, Greenfield S, Ware J. The functioning and well-being of depressed patients: results from the Medical Outcomes Study. Journal of the American Medical Association. 1989; 262:914–919. [PubMed: 2754791]
8. Broadhead WE, Blazer DG, George LK, Tse CK. Depression, disability days, and days lost from work in a prospective epidemiologic survey. Journal of the American Medical Association. 1990; 264:2524–2528. [PubMed: 2146410]
9. Judd LL, Paulus MP, Wells KB, Rapaport MN. Socioeconomic burden of subsyndromal depressive symptoms and major depression in a sample of the general population. American Journal of Psychiatry. 1996; 153:1411–1417. [PubMed: 8890673]
10. Tsuang MT, Woolson RF. Excess mortality in schizophrenia and affective disorders. Archives of General Psychiatry. 1978; 35:1181–1185. [PubMed: 697536]
11. Berglund M, Nilsson K. Mortality in severe depression: a prospective study including 103 suicides. Acta Psychiatrica Scandinavica. 1987; 76:372–380. [PubMed: 3425363]
12. Black DW, Winokur G, Nasrallah A. Is death from natural causes still excessive in psychiatric patients? Journal of Nervous and Mental Disease. 1987; 175:674–680. [PubMed: 3681278]

13. Zilber N, Schufman N, Lerner Y. Mortality among psychiatric patients - the groups at risk. Acta Psychiatrica Scandinavica. 1989; 79:248–256. [PubMed: 2711851]
14. Greenberg PE, Stiglin LE, Finkelstein SN, Berndt ER. The economic burden of depression in 1990. Journal of Clinical Psychiatry. 1993; 54:405–418. [PubMed: 8270583]
15. Murray CJL, Lopez AD. Evidence-based health policy: lessons from the Global Burden of Disease Study. Science. 1996; 274:740–743. [PubMed: 8966556]
16. Sullivan PF, Neale MC, Kendler KS. Genetic epidemiology of major depression: review and meta-analysis. American Journal of Psychiatry. 2000; 157:1552–1562. [PubMed: 11007705]
17. Altshuler D, Daly M. Guilt beyond a reasonable doubt. Nat Genet. 2007; 39(7):813–5. [PubMed: 17597768]
18. Psychiatric GWAS Consortium. A framework for interpreting genomewide association studies of psychiatric disorders. Submitted.
19. Manolio TA, Rodriguez LL, Brooks L, Abecasis G, Ballinger D, Daly M, Donnelly P, Faraone SV, Frazer K, Gabriel S, Gejman P, Guttmacher A, Harris EL, Insel T, Kelsoe JR, Lander E, McCowin N, Mailman MD, Nabel E, Ostell J, Pugh E, Sherry S, Sullivan PF, Thompson JF, Warram J, Wholley D, Milos PM, Collins FS. New models of collaboration in genome-wide association studies: the Genetic Association Information Network. Nat Genet. 2007; 39(9):1045–1051. [PubMed: 17728769]
20. Mailman MD, Feolo M, Jin Y, Kimura M, Tryka K, Bagoutdinov R, Hao L, Kiang A, Paschall J, Phan L, Popova N, Pretel S, Ziyabari L, Lee M, Shao Y, Wang ZY, Sirotkin K, Ward M, Kholodov M, Zbicz K, Beck J, Kimelman M, Shevelev S, Preuss D, Yaschenko E, Graeff A, Ostell J, Sherry ST. The NCBI dbGaP database of genotypes and phenotypes. Nat Genet. 2007; 39(10):1181–6. [PubMed: 17898773]
21. Chanock SJ, Manolio T, Boehnke M, Boerwinkle E, Hunter DJ, Thomas G, Hirschhorn JN, Abecasis G, Altshuler D, Bailey-Wilson JE, Brooks LD, Cardon LR, Daly M, Donnelly P, Fraumeni JF Jr. Freimer NB, Gerhard DS, Gunter C, Guttmacher AE, Guyer MS, Harris EL, Hoh J, Hoover R, Kong CA, Merikangas KR, Morton CC, Palmer LJ, Phimister EG, Rice JP, Roberts J, Rotimi C, Tucker MA, Vogan KJ, Wacholder S, Wijsman EM, Winn DM, Collins FS. Replicating genotype-phenotype associations. Nature. 2007; 447(7145):655–60. [PubMed: 17554299]
22. Penninx B, Beekman A, Smit J. The Netherlands Study of Depression and Anxiety (NESDA): Rationales, Objectives and Methods. International Journal of Methods in Psychiatric Research. 2008; 17:121–40. [PubMed: 18763692]
23. Boomsma DI, de Geus EJ, Vink JM, Stubbe JH, Distel MA, Hottenga JJ, Posthuma D, van Beijsterveldt TC, Hudziak JJ, Bartels M, Willemsen G. Netherlands Twin Register: from twins to twin families. Twin Res Hum Genet. 2006; 9(6):849–57. [PubMed: 17254420]
24. Boomsma DI, Willemsen G, Sullivan PF, Heutnik P, Meijer P, Sondervan D, Kluft C, Smit G, Nolen WA, Zitman FG, Smit JH, Hoogendijk WJ, van Dyck R, de Geus EJC, Penninx BWJH. Genome-wide association of major depression: Description of samples for the GAIN major depressive disorder study: NTR and NESDA Biobank Projects. In press.
25. Bijl RV, van Zessen G, Ravelli A, de Rijk C, Langendoen Y. The Netherlands Mental Health Survey and Incidence Study (NEMESIS): objectives and design. Soc Psychiatry Psychiatr Epidemiol. 1998; 33(12):581–6. [PubMed: 9857790]
26. Landman-Peeters KM, Hartman CA, van der Pompe G, den Boer JA, Minderaa RB, Ormel J. Gender differences in the relation between social support, problems in parent-offspring communication, and depression and anxiety. Soc Sci Med. 2005; 60(11):2549–59. [PubMed: 15814180]
27. World Health Organization. Composite International Diagnostic Interview (CIDI), Version 2.1. World Health Organization; Geneva, Switzerland: 1997.
28. Boomsma DI, Beem AL, van den Berg M, Dolan CV, Koopmans JR, Vink JM, de Geus EJ, Slagboom PE. Netherlands twin family study of anxious depression (NETSAD). Twin Res. 2000; 3(4):323–34. [PubMed: 11463154]
29. Kessler RC, Barker PR, Colpe LJ, Epstein JF, Gfroerer JC, Hiripi E, Howes MJ, Normand SL, Manderscheid RW, Walters EE, Zaslavsky AM. Screening for serious mental illness in the general population. Arch Gen Psychiatry. 2003; 60(2):184–9. [PubMed: 12578436]

30. Rush AJ, Gullion CM, Basco MR, Jarrett RB, Trivedi MH. The Inventory of Depressive Symptomatology (IDS): psychometric properties. Psychol Med. 1996; 26(3):477–86. [PubMed: 8733206]

31. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Y, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Sun W, Wang H, Wang Y, Xiong X, Xu L, Waye MM, Tsui SK, Xue H, Wong JT, Galver LM, Fan JB, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS, Roumy S, Sallee C, Verner A, Hudson TJ, Kwok PY, Cai D, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui LC, Mak W, Song YQ, Tam PK, Nakamura Y, Kawaguchi T, Kitamoto T, Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T, Tsunoda T, Deloukas P, Bird CP, Delgado M, Dermitzakis ET, Gwilliam R, Hunt S, Morrison J, Powell D, Stranger BE, Whittaker P, Bentley DR, Daly MJ, de Bakker PI, Barrett J, Chretien YR, Maller J, McCarroll S, Patterson N, Pe'er I, Price A, Purcell S, Richter DJ, Sabeti P, Saxena R, Schaffner SF, Sham PC, Varilly P, Stein LD, Krishnan L, Smith AV, Tello-Ruiz MK, Thorisson GA, Chakravarti A, Chen PE, Cutler DJ, Kashuk CS, Lin S, Abecasis GR, Guan W, Li Y, Munro HM, Qin ZS, Thomas DJ, McVean G, Auton A, Bottolo L, Cardin N, Eyheramendy S, Freeman C, Marchini J, Myers S, Spencer C, Stephens M, Donnelly P, Cardon LR, Clarke G, Evans DM, Morris AP, Weir BS, Mullikin JC, Sherry ST, Feolo M, Skol A, Zhang H, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo CA, Ajayi I, Aniagwu T, Marshall PA, Nkwodimmah C, Royal CD, Leppert MF, Dixon M, Peiffer A, Qiu R, Kent A, Kato K, Niikawa N, Adewole IF, Knoppers BM, Foster MW, Clayton EW, Watkin J, Muzny D, Nazareth L, Sodergren E, Weinstock GM, Yakub I, Birren BW, Wilson RK, Fulton LL, Rogers J, Burton J, Carter NP, Clee CM, Griffiths M, Jones MC, McLay K, Plumb RW, Ross MT, Sims SK, Willey DL, Chen Z, Han H, Kang L, Godbout M, Wallenburg JC, L'Archeveque P, Bellemare G, Saeki K, An D, Fu H, Li Q, Wang Z, Wang R, Holden AL, Brooks LD, McEwen JE, Guyer MS, Wang VO, Peterson JL, Shi M, Spiegel J, Sung LM, Zacharia LF, Collins FS, Kennedy K, Jamieson R, Stewart J. A second generation human haplotype map of over 3.1 million SNPs. Nature. 2007; 449(7164):851–61. [PubMed: 17943122]

32. Hinds DA, Stuve LL, Nilsen GB, Halperin E, Eskin E, Ballinger DG, Frazer KA, Cox DR. Whole-genome patterns of common DNA variation in three human populations. Science. 2005; 307(5712):1072–9. [PubMed: 15718463]

33. Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. Am J Hum Genet. 2004; 74(1):106–20. [PubMed: 14681826]

34. Altshuler D, Brooks LD, Chakravarti A, Collins FS, Daly MJ, Donnelly P. A haplotype map of the human genome. Nature. 2005; 437(7063):1299–320. [PubMed: 16255080]

35. Bierut LJ, Madden PA, Breslau N, Johnson EO, Hatsukami D, Pomerleau OF, Swan GE, Rutter J, Bertelsen S, Fox L, Fugman D, Goate AM, Hinrichs AL, Konvicka K, Martin NG, Montgomery GW, Saccone NL, Saccone SF, Wang JC, Chase GA, Rice JP, Ballinger DG. Novel genes identified in a high-density genome wide association study for nicotine dependence. Hum Mol Genet. 2007; 16(1):24–35. [PubMed: 17158188]

36. Sullivan PF, Lin D, Tzeng JY, van den Oord EJCG, Perkins D, Stroup TS, Wagner M, Lee S, Wright FA, Zou F, Liu W, Downing AM, Lieberman JA, Close SL. Genomewide association for schizophrenia in the CATIE study: Results of Stage 1. Molecular Psychiatry. 2008; 13:570–84. [PubMed: 18347602]

37. Hemminger BM, Saelim B, Sullivan PF. TAMAL: An integrated approach to choosing SNPs for genetic studies of human complex traits. Bioinformatics. 2006; 22:626–7. [PubMed: 16418238]

38. Wittke-Thompson JK, Pluzhnikov A, Cox NJ. Rational inferences about departures from Hardy-Weinberg equilibrium. Am J Hum Genet. 2005; 76(6):967–86. [PubMed: 15834813]

39. Wigginton JE, Cutler DJ, Abecasis GR. A note on exact tests of Hardy-Weinberg equilibrium. Am J Hum Genet. 2005; 76(5):887–3. [PubMed: 15789306]

40. Balding DJ. A tutorial on statistical methods for population association studies. Nat Rev Genet. 2006; 7(10):781–91. [PubMed: 16983374]
41. Sasieni PD. From genotypes to genes: doubling the sample size. Biometrics. 1997; 53:1253–1261. [PubMed: 9423247]
42. Fisher, RA. Statistical Methods for Reesearch Workers. 11th edition. Oliver and Boyd; London: 1950.
43. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M, Bender D, Maller J, de Bakker P, Daly M, Sham P. PLINK: a toolset for whole-genome association and population-based linkage analysis. American Journal of Human Genetics. 2007; 81:559–75. [PubMed: 17701901]
44. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics. 2006; 38:904–9. [PubMed: 16862161]
45. Storey JD. The positive false discovery rate: a Bayesian interpretation and the q-value. Annals of Statistics. 2003; 31:2013–35.
46. Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc Natl Acad Sci U S A. 2003; 100(16):9440–5. [PubMed: 12883005]
47. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society (Series B). 1995; 57:289–300.
48. Brown BW, Russell K. Methods of correcting for multiple testing: operating characteristics. Stat Med. 1997; 16(22):2511–28. [PubMed: 9403953]
49. Fernando RL, Nettleton D, Southey BR, Dekkers JC, Rothschild MF, Soller M. Controlling the proportion of false positives in multiple dependent tests. Genetics. 2004; 166(1):611–9. [PubMed: 15020448]
50. van den Oord EJ, Sullivan PF. A framework for controlling false discovery rates and minimizing the amount of genotyping in the search for disease mutations. Hum Hered. 2003; 56(4):188–99. [PubMed: 15031620]
51. Tsai CA, Hsueh HM, Chen JJ. Estimation of false discovery rates in multiple testing: application to gene microarray data. Biometrics. 2003; 59(4):1071–81. [PubMed: 14969487]
52. van den Oord EJ. Controlling false discoveries in candidate gene studies. Mol Psychiatry. 2005; 10(3):230–1. [PubMed: 15738930]
53. Sabatti C, Service S, Freimer N. False discovery rate in linkage and association genome screens for complex disorders. Genetics. 2003; 164(2):829–33. [PubMed: 12807801]
54. Meinhausen N, Rice J. Estimating the proportion of false null hypotheses among a large number of independently tested hypotheses. Annals of Statistics. 2006; 34:373–93.
55. van den Oord EJ, Sullivan PF. False discoveries and models for gene discovery. Trends Genet. 2003; 19(10):537–42. [PubMed: 14550627]
56. Lin DY, Hu Y, Huang BE. Simple and efficient analysis of disease association with missing genotype data. American Journal of Human Genetics. 2008; 82:444–452. [PubMed: 18252224]
57. Gauderman WJ. Sample size requirements for association studies of gene-gene interaction. Am J Epidemiol. 2002; 155(5):478–84. [PubMed: 11867360]
58. Gauderman WJ. Sample size requirements for matched case-control studies of gene-environment interaction. Stat Med. 2002; 21(1):35–50. [PubMed: 11782049]
59. SAS Institute Inc. SAS/STAT® Software: Version 9. SAS Institute, Inc.; Cary, NC: 2004.
60. R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; Vienna, Austria: 2007.
61. Lin DY, Zeng D, Millikan R. Maximum likelihood estimation of haplotype effects and haplotype-environment interactions in association studies. Genet Epidemiol. 2005; 29(4):299–312. [PubMed: 16240443]
62. Zeng D, Lin DY, Avery CL, North KE, Bray MS. Efficient semiparametric estimation of haplotype-disease associations in case-cohort and nested case-control studies. Biostatistics. 2006; 7(3):486–502. [PubMed: 16500923]
63. Huang B, Amos C, Lin D. Detecting haplotype effects in genomewide association studies. Genetic Epidemiology. 2007; 31:803–12. [PubMed: 17549762]

64. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005; 21:263–5. [PubMed: 15297300]

65. SAS Institute Inc. JMP User's Guide (Version 6). SAS Institute, Inc.; Cary, NC: 2005.

66. Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Edgar R, Federhen S, Geer LY, Helmberg W, Kapustin Y, Kenton DL, Khovayko O, Lipman DJ, Madden TL, Maglott DR, Ostell J, Pruitt KD, Schuler GD, Schriml LM, Sequeira E, Sherry ST, Sirotkin K, Souvorov A, Starchenko G, Suzek TO, Tatusov R, Tatusova TA, Wagner L, Yaschenko E. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2006; 34(Database issue):D173–80. [PubMed: 16381840]

67. Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, Diekhans M, Furey TS, Harte RA, Hsu F, Hillman-Jackson J, Kuhn RM, Pedersen JS, Pohl A, Raney BJ, Rosenbloom KR, Siepel A, Smith KE, Sugnet CW, Sultan-Qurraie A, Thomas DJ, Trumbower H, Weber RJ, Weirauch M, Zweig AS, Haussler D, Kent WJ. The UCSC Genome Browser Database: update 2006. Nucleic Acids Res. 2006; 34(Database issue):D590–8. [PubMed: 16381938]

68. Blaschke RJ, Rappold G. The pseudoautosomal regions, SHOX and disease. Curr Opin Genet Dev. 2006; 16(3):233–9. [PubMed: 16650979]

69. Devlin B, Roeder K. Genomic control for association studies. Biometrics. 1999; 55:997–1004. [PubMed: 11315092]

70. Baum AE, Akula N, Cabanero M, Cardona I, Corona W, Klemens B, Schulze TG, Cichon S, Rietschel M, Nothen MM, Georgi A, Schumacher J, Schwarz M, Abou Jamra R, Hofels S, Propping P, Satagopan J, Detera-Wadleigh SD, Hardy J, McMahon FJ. A genome-wide association study implicates diacylglycerol kinase eta (DGKH) and several other genes in the etiology of bipolar disorder. Mol Psychiatry. 2007

71. Phillips GR, Huang JK, Wang Y, Tanaka H, Shapiro L, Zhang W, Shan WS, Arndt K, Frank M, Gordon RE, Gawinowicz MA, Zhao Y, Colman DR. The presynaptic particle web: ultrastructure, composition, dissolution, and reconstitution. Neuron. 2001; 32(1):63–77. [PubMed: 11604139]

72. Schildkraut JJ. The catecholamine hypothesis of affective disorders: a review of supporting evidence. Am J Psychiatry. 1965; 122(5):509–22. [PubMed: 5319766]

73. Wang X, Kibschull M, Laue MM, Lichte B, Petrasch-Parwez E, Kilimann MW. Aczonin, a 550-kD putative scaffolding protein of presynaptic active zones, shares homology regions with Rim and Bassoon and binds profilin. J Cell Biol. 1999; 147(1):151–62. [PubMed: 10508862]

74. Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie X, Byrne EH, McCarroll SA, Gaudet R, Schaffner SF, Lander ES, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Y, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Sun W, Wang H, Wang Y, Xiong X, Xu L, Waye MM, Tsui SK, Xue H, Wong JT, Galver LM, Fan JB, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS, Roumy S, Sallee C, Verner A, Hudson TJ, Kwok PY, Cai D, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui LC, Mak W, Song YQ, Tam PK, Nakamura Y, Kawaguchi T, Kitamoto T, Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T, Tsunoda T, Deloukas P, Bird CP, Delgado M, Dermitzakis ET, Gwilliam R, Hunt S, Morrison J, Powell D, Stranger BE, Whittaker P, Bentley DR, Daly MJ, de Bakker PI, Barrett J, Chretien YR, Maller J, McCarroll S, Patterson N, Pe'er I, Price A, Purcell S, Richter DJ, Sabeti P, Saxena R, Sham PC, Stein LD, Krishnan L, Smith AV, Tello-Ruiz MK, Thorisson GA, Chakravarti A, Chen PE, Cutler DJ, Kashuk CS, Lin S, Abecasis GR, Guan W, Li Y, Munro HM, Qin ZS, Thomas DJ, McVean G, Auton A, Bottolo L, Cardin N, Eyheramendy S, Freeman C, Marchini J, Myers S, Spencer C, Stephens M, Donnelly P, Cardon LR, Clarke G, Evans DM, Morris AP, Weir BS, Johnson TA, Mullikin JC, Sherry ST, Feolo M, Skol A, Zhang H, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo CA, Ajayi I, Aniagwu T, Marshall PA, Nkwodimmah C, Royal CD, Leppert MF, Dixon M, Peiffer A, Qiu R, Kent A, Kato K, Niikawa N, Adewole IF, Knoppers BM, Foster MW, Clayton EW, Watkin J, Muzny D, Nazareth L, Sodergren E, Weinstock GM, Yakub I, Birren BW, Wilson RK, Fulton LL, Rogers J, Burton J, Carter NP, Clee CM, Griffiths M, Jones MC, McLay K, Plumb RW, Ross

MT, Sims SK, Willey DL, Chen Z, Han H, Kang L, Godbout M, Wallenburg JC, L'Archeveque P, Bellemare G, Saeki K, An D, Fu H, Li Q, Wang Z, Wang R, Holden AL, Brooks LD, McEwen JE, Guyer MS, Wang VO, Peterson JL, Shi M, Spiegel J, Sung LM, Zacharia LF, Collins FS, Kennedy K, Jamieson R, Stewart J. Genome-wide detection and characterization of positive selection in human populations. Nature. 2007; 449(7164):913–8. [PubMed: 17943131]

75. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C. Detection of large-scale variation in the human genome. Nat Genet. 2004; 36(9):949–51. [PubMed: 15286789]
76. Pinto D, Marshall C, Feuk L, Scherer SW. Copy-number variation in control population cohorts. Hum Mol Genet. 2007; 16(Spec No. 2):R168–73. [PubMed: 17911159]
77. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF, Hakonarson H, Bucan M. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res. 2007; 17(11):1665–74. [PubMed: 17921354]
78. Levinson DF, Zubenko GS, Crowe RR, DePaulo RJ, Scheftner WS, Weissman MM, Holmans P, Zubenko WN, Boutelle S, Murphy-Eberenz K, MacKinnon D, McInnis MG, Marta DH, Adams P, Sassoon S, Knowles JA, Thomas J, Chellis J. Genetics of recurrent early-onset depression (GenRED): design and preliminary clinical characteristics of a repository sample for genetic linkage studies. Am J Med Genet B Neuropsychiatr Genet. 2003; 119(1):118–30. [PubMed: 12707949]
79. Sun L, Bull S. Reduction of selection bias in genomewide genetic studies by resampling. Genetic Epidemiology. 2005; 28:352–67. [PubMed: 15761913]
80. de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D. Efficiency and power in genetic association studies. Nat Genet. 2005; 37(11):1217–23. [PubMed: 16244653]
81. Lopez-Leon S, Janssens AC, Gonzalez-Zuloeta Ladd AM, Del-Favero J, Claes SJ, Oostra BA, van Duijn CM. Meta-analyses of genetic studies on major depressive disorder. Mol Psychiatry. 2007
82. Frayling TM. Genome-wide association studies provide new insights into type 2 diabetes aetiology. Nature Reviews Genetics. 2007; 8(9):657–62.
83. Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU, Prokunina-Olsson L, Ding CJ, Swift AJ, Narisu N, Hu T, Pruim R, Xiao R, Li XY, Conneely KN, Riebow NL, Sprau AG, Tong M, White PP, Hetrick KN, Barnhart MW, Bark CW, Goldstein JL, Watkins L, Xiang F, Saramies J, Buchanan TA, Watanabe RM, Valle TT, Kinnunen L, Abecasis GR, Pugh EW, Doheny KF, Bergman RN, Tuomilehto J, Collins FS, Boehnke M. A Genome-Wide Association Study of Type 2 Diabetes in Finns Detects Multiple Susceptibility Variants. Science. 2007; 316:1341–5. [PubMed: 17463248]
84. Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PI, Chen H, Roix JJ, Kathiresan S, Hirschhorn JN, Daly MJ, Hughes TE, Groop L, Altshuler D, Almgren P, Florez JC, Meyer J, Ardlie K, Bengtsson K, Isomaa B, Lettre G, Lindblad U, Lyon HN, Melander O, Newton-Cheh C, Nilsson P, Orho-Melander M, Rastam L, Speliotes EK, Taskinen MR, Tuomi T, Guiducci C, Berglund A, Carlson J, Gianniny L, Hackett R, Hall L, Holmkvist J, Laurila E, Sjogren M, Sterner M, Surti A, Svensson M, Tewhey R, Blumenstiel B, Parkin M, Defelice M, Barry R, Brodeur W, Camarata J, Chia N, Fava M, Gibbons J, Handsaker B, Healy C, Nguyen K, Gates C, Sougnez C, Gage D, Nizzari M, Gabriel SB, Chirn GW, Ma Q, Parikh H, Richardson D, Ricke D, Purcell S. Genome-Wide Association Analysis Identifies Loci for Type 2 Diabetes and Triglyceride Levels. Science. 2007; 316:1331–6. [PubMed: 17463246]
85. Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T, de Bakker PI, Abecasis GR, Almgren P, Andersen G, Ardlie K, Bostrom KB, Bergman RN, Bonnycastle LL, Borch-Johnsen K, Burtt NP, Chen H, Chines PS, Daly MJ, Deodhar P, Ding CJ, Doney AS, Duren WL, Elliott KS, Erdos MR, Frayling TM, Freathy RM, Gianniny L, Grallert H, Grarup N, Groves CJ, Guiducci C, Hansen T, Herder C, Hitman GA, Hughes TE, Isomaa B, Jackson AU, Jorgensen T, Kong A, Kubalanza K, Kuruvilla FG, Kuusisto J, Langenberg C, Lango H, Lauritzen T, Li Y, Lindgren CM, Lyssenko V, Marvelle AF, Meisinger C, Midthjell K, Mohlke KL, Morken MA, Morris AD, Narisu N, Nilsson P, Owen KR, Palmer CN, Payne F, Perry JR, Pettersen E, Platou C, Prokopenko I, Qi L, Qin L, Rayner NW, Rees M, Roix JJ, Sandbaek A, Shields B, Sjogren M, Steinthorsdottir V, Stringham HM, Swift AJ, Thorleifsson G, Thorsteinsdottir U, Timpson NJ, Tuomi T, Tuomilehto J, Walker M, Watanabe RM, Weedon MN, Willer CJ, Illig T, Hveem K, Hu

FB, Laakso M, Stefansson K, Pedersen O, Wareham NJ, Barroso I, Hattersley AT, Collins FS, Groop L, McCarthy MI, Boehnke M, Altshuler D. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nat Genet. 2008; 40(5):638–45. [PubMed: 18372903]

86. Sudhof TC. Neurotransmitter release. Handb Exp Pharmacol. 2008; (184):1–21.
87. Shildkraut JJ. The catecholamine hypothesis of affective disorders: a review of the supporting evidence. American Journal of Psychiatry. 1965; 122:509–522. [PubMed: 5319766]
88. Garcia J, Gerber SH, Sugita S, Sudhof TC, Rizo J. A conformational switch in the Piccolo C2A domain regulated by alternative splicing. Nat Struct Mol Biol. 2004; 11(1):45–53. [PubMed: 14718922]
89. Gerber SH, Garcia J, Rizo J, Sudhof TC. An unusual C(2)-domain in the active-zone protein piccolo: implications for Ca(2+) regulation of neurotransmitter release. Embo J. 2001; 20(7): 1605–19. [PubMed: 11285225]
90. Nabi R, Zhong H, Serajee FJ, Huq AH. No association between single nucleotide polymorphisms in DLX6 and Piccolo genes at 7q21-q22 and autism. Am J Med Genet B Neuropsychiatr Genet. 2003; 119B(1):98–101. [PubMed: 12707945]
91. Zollner S, Pritchard JK. Overcoming the winner's curse: estimating penetrance parameters from case-control data. Am J Hum Genet. 2007; 80(4):605–15. [PubMed: 17357068]
92. Ghosh A, Zou F, Wright FA. Estimating odds ratios in genome scans: an approximate conditional likelihood approach. Am J Hum Genet. 2008; 82(5):1064–74. [PubMed: 18423522]
93. Patten SB. Selection bias in studies of major depression using clinical subjects. J Clin Epidemiol. 2000; 53(4):351–7. [PubMed: 10785565]
94. Galbaud du Fort G, Newman SC, Bland RC. Psychiatric comorbidity and treatment seeking. Sources of selection bias in the study of clinical populations. J Nerv Ment Dis. 1993; 181(8):467–74. [PubMed: 8360638]
95. Berkson J. Limitations of the application of fourfold table analysis to hospital data. Biometics Bulletin. 1946; 2:47–53.
96. Sullivan PF, Joyce PR. Effects of exclusion criteria in depression treatment studies. Journal of Affective Disorders. 1994; 32:21–26. [PubMed: 7798463]
97. Sullivan PF, Wells JE, Joyce PR, Bushnell JA, Mulder RT, Oakley-Browne MA. Family history of depression in clinic and community samples. Journal of Affective Disorders. 1996; 40:159–168. [PubMed: 8897115]
98. Kendler KS, Neale MC, Kessler RC, Heath AC, Eaves LJ. The lifetime history of major depression in women: reliability of diagnosis and heritability. Archives of General Psychiatry. 1993; 50:863–870. [PubMed: 8215812]
99. McGuffin P, Katz R, Watkins S, Rutherford J. A hospital-based twin register of the heritability of DSM-IV unipolar depression. Archives of General Psychiatry. 1996; 53:129–136. [PubMed: 8629888]
100. Shaw CJ, Lupski JR. Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease. Hum Mol Genet. 2004; 13(Spec No 1):R57–64. [PubMed: 14764619]
101. Konneker T, Barnes T, Furberg H, Losh M, Bulik CM, Sullivan PF. A searchable database of genetic evidence for psychiatric disorders. American Journal of Medical Genetics (Neuropsychiatric Genetics). 2008; 147:671–5.
102. WTCCC. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007; 447(7145):661–78. [PubMed: 17554300]
103. van den Oord EJ, Kuo PH, Hartmann AM, Webb BT, Moller HJ, Hettema JM, Giegling I, Bukszar J, Rujescu D. Genomewide association analysis followed by a replication study implicates a novel candidate gene for neuroticism. Arch Gen Psychiatry. 2008; 65(9):1062–71. [PubMed: 18762592]
104. O'Donovan M, Craddock N, Norton N, Williams H, Peirce T, Moskvina V, Nikolov I, Hamshere M, Carroll L, Georgieva L, Dwyer S, Holmans P, Marchini J, Spencer C, Howie B, Leung H-T, Hartmann A, Möller H-J, Morris D, Shi Y, Feng G, Hoffmann P, Propping P, Vasilescu C, Maier W, Rieschel M, Zammit S, Schumacher J, Quinn E, Schulze T, Williams N, Giegling I, Iwata N,

Ikeda M, Darvasi A, Shifman S, He L, Duan J, Sanders A, Levinson D, Gejman P, Cichon S, Nöthen M, Gill M, Corvin A, Rujescu D, Kirov G, Owen M. Identification of novel schizophrenia loci by genome-wide association and follow-up. Nature Genetics. 2008

105. Ferreira M, O'Donovan M, Meng Y, Jones I, Ruderfer D, Jones L, Fan J, Kirov G, Perlis R, Green E, Smoller J, Grozeva D, Stone J, Nikolov I, Chambert K, Hamshere M, Nimgaonkar V, Moskvina V, Thase M, Caesar S, Sachs G, Franklin J, Gordon-Smith K, Ardlie K, Gabriel S, Fraser C, Blumenstiel B, Defelice M, Breen G, Gill M, Morris D, Elkin A, Muir W, McGhee K, Williamson R, MacIntyre D, McLean A, St Clair D, VanBeck M, Pereira A, Kandaswamy R, McQuillin A, Collier D, Bass N, Young A, Lawrence J, Ferrier I, Anjorin A, Farmer A, Curtis D, Scolnick E, McGuffin P, Daly M, Corvin A, Holmans P, Blackwood D, Gurling H, Owen M, Purcell S, Sklar P, Craddock N. Collaborative genome-wide association analysis of 10,596 individuals supports a role for Ankyrin-G (*ANK3*) and the alpha-1C subunit of the L-type voltage-gated calcium channel (*CACNA1C*) in bipolar disorder. Nature Genetics. 2008

106. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi M, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M. Large-scale copy number polymorphism in the human genome. Science. 2004; 305(5683):525–8. [PubMed: 15273396]

107. Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, Fossdal R, Saemundsen E, Stefansson H, Ferreira MA, Green T, Platt OS, Ruderfer DM, Walsh CA, Altshuler D, Chakravarti A, Tanzi RE, Stefansson K, Santangelo SL, Gusella JF, Sklar P, Wu BL, Daly MJ. Association between microdeletion and microduplication at 16p11.2 and autism. New England Journal of Medicine. 2008; 358(7):667–75. [PubMed: 18184952]

108. Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM, Nord AS, Kusenda M, Malhotra D, Bhandari A, Stray SM, Rippey CF, Roccanova P, Makarov V, Lakshmi B, Findling RL, Sikich L, Stromberg T, Merriman B, Gogtay N, Butler P, Eckstrand K, Noory L, Gochman P, Long R, Chen Z, Davis S, Baker C, Eichler EE, Meltzer PS, Nelson SF, Singleton AB, Lee MK, Rapoport JL, King MC, Sebat J. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. Science. 2008; 320:539–43. [PubMed: 18369103]

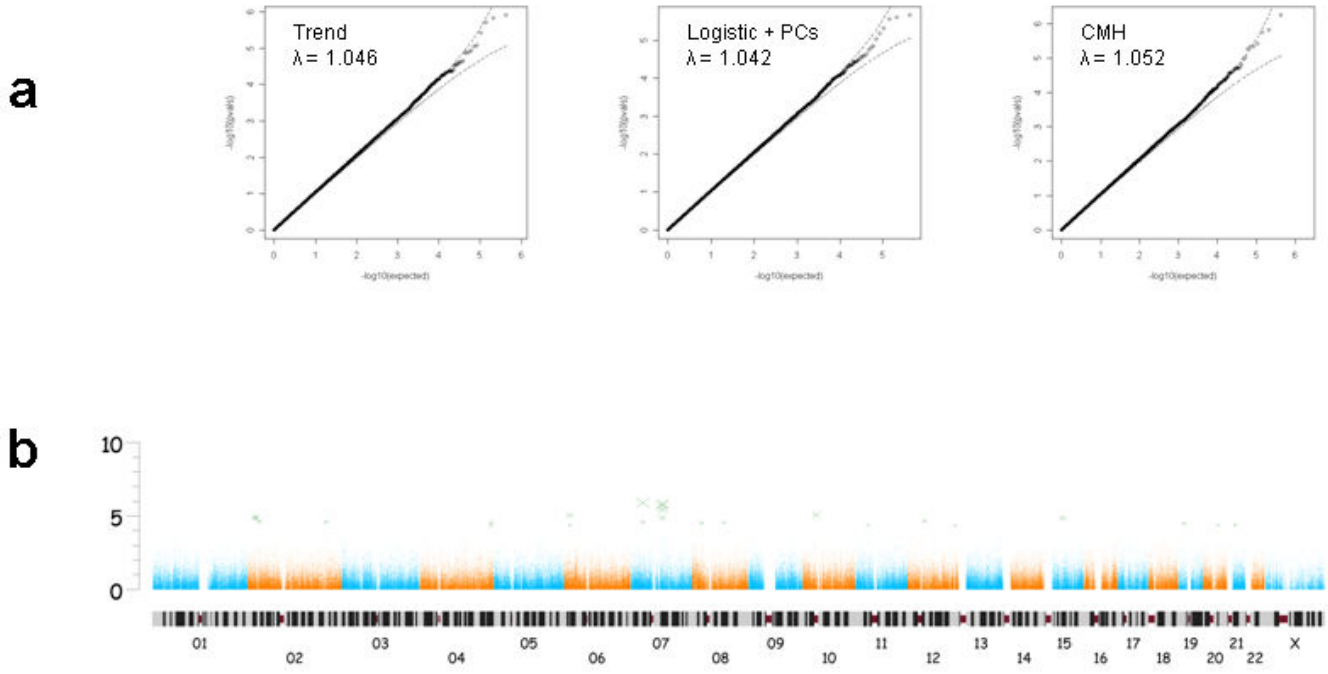

Figure 1.

GWAS results for MDD in cases versus controls. Panel a shows the quantile-quantile plots and lambda estimates for the primary analysis using the Cochran-Armitage trend test and confirmatory analyses using logistic regressions and Cochran-Mantel-Haenszel stratified tests. The dashed lines show the expected 95% probability interval for ordered p-values, and the circles show the observed versus expected values for all SNPs. The λ values are the median X^2 from all association tests divided by the expected value under the null hypothesis of no association. If λ is large (e.g., >1.2), there is evidence that the observed test statistics deviate from the expected. This could be due to true associations but is more likely due to a systematic bias (e.g., population stratification effects). The λ values in Panel a are not consistent with the presence of stematic biases in the results. Panel b depicts $-\log_{10}(p)$ by genomic location for chr1-chr22 plus chrX. Odd-numbered chromosomes are in light blue and even-numbered chromosomes in orange. The 25 smallest p-values are shown with green crosses.

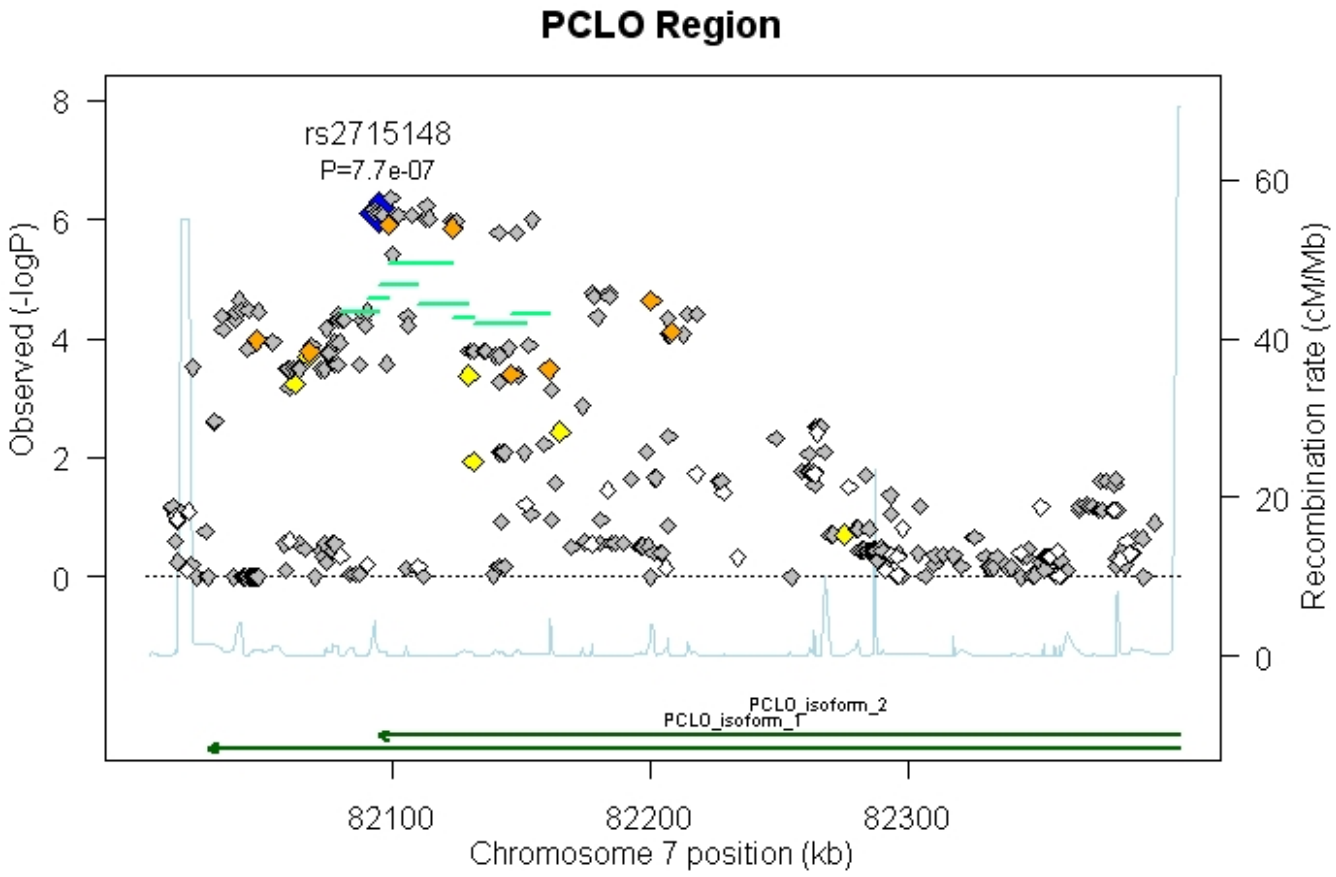

Figure 2.

Plot of the *PCLO* region (NCBI build 35, UCSC hg17, chr7:82,000,000-82,500,000). P-values in this figure are all from SNPMstat. The x-axis is chromosomal position, the left y-axis $-\log_{10}(p)$ for genotyped SNPs (colored diamonds) and imputed SNPs (grey diamonds), and the right y-axis the recombination rate from the HapMap EUR panel (light blue curve). The color of the genotyped SNPs corresponds to LD with the SNP with smallest p-value (rs2715148): red $0.8 \leq r^2 \leq 1.0$, orange $0.5 \leq r^2 < 0.8$, yellow $0.2 \leq r^2 < 0.5$, and white $r^2 < 0.2$. The significant and extent of all 3-SNP haplotypes with $p<0.0001$ in this region are colored light green. The transcripts for two *PCLO* isoforms are shown in dark green at the bottom. Graph adapted from an R function by the Broad DGI group.

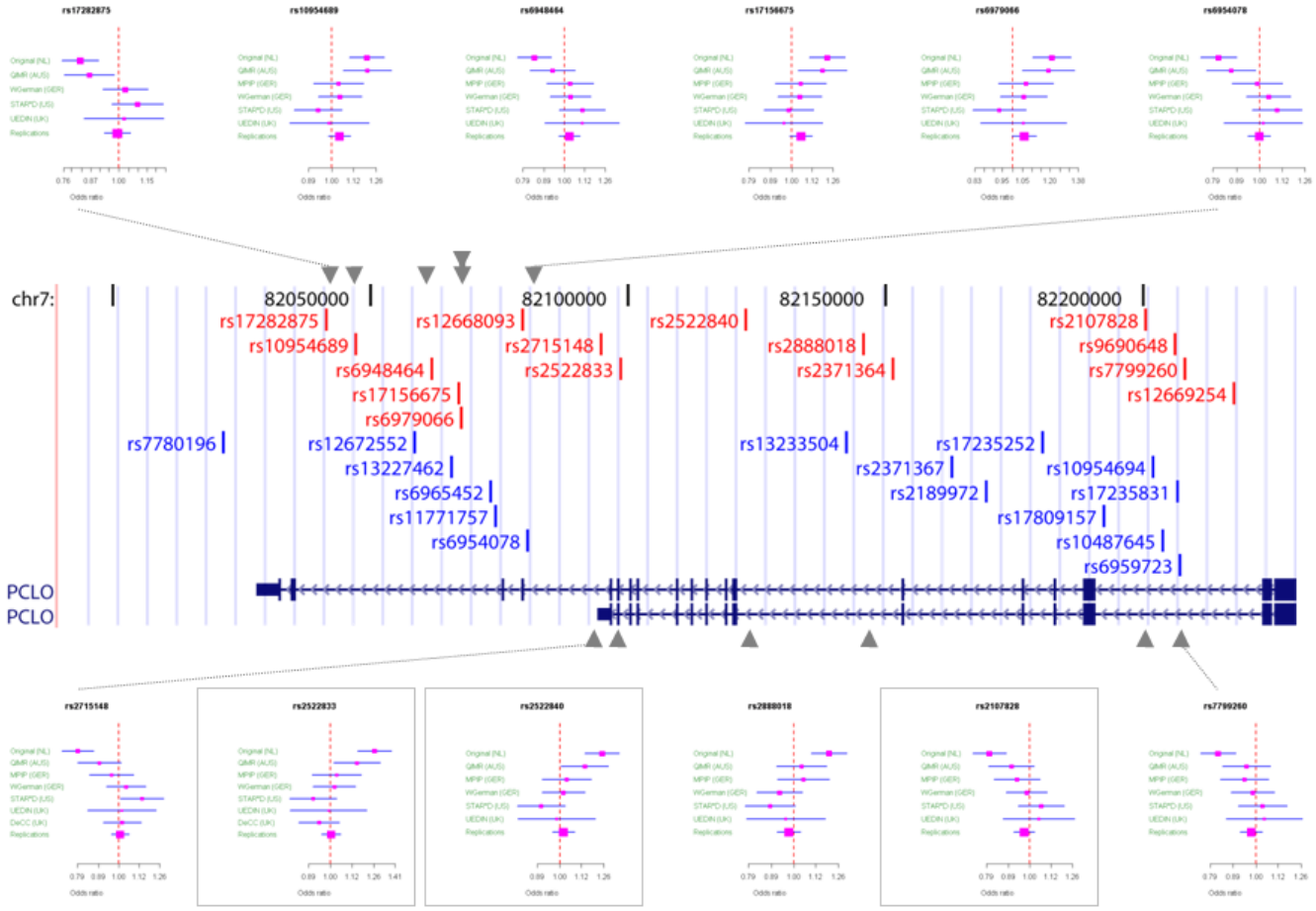

Figure 3.

PCLO region replication results for MDD showing genomic context and forest plots for the top 12 SNPs in the original sample. The backbone of the graph is the region of *PCLO* targeted for follow-up. SNP locations are given by the grey triangles. There are 12 forest plots for the SNPs with p<0.001 in the original sample. Each forest plot is for one SNP and shows the odds ratio (square) and 95% confidence intervals (horizontal line) for a particular sample with the area of the square proportional to sample size.

Table 1

Descriptive data for cases with MDD and controls at low liability for MDD included in the GWAS.

Descriptor	Cases	Controls	Test
Number of subjects genotyped	1,738	1,802	-
Mean age in years (SD)	42.6 (12.6)	45.1 (14.1)	$F_{1,3538}$=31.1, p<0.001
Female, %	69.6	62.0	χ_1^2 = 22.5, p<0.001
Educational level (% low/middle/high)	7.8 / 62.0 / 32.2	5.7 / 56.3 / 38.1	χ_2^2 = 16.3, p<0.001
Partner status, % with partner	68.9	87.0	χ_1^2 = 167.2, p<0.001
Smoking (current), %	42.0	20.2	χ_1^2 = 194.5, p<0.001
Mean neuroticism (NEO, SD)	39.3 (8.0)	28.2 (5.5)	$F_{1,2920}$=1831, p<0.001
MDD, age of onset in years (SD) Early age of onset (< 30 years), %	27.7 (12.4) 57.3	-	
Family history of depression, %	85.5	-	
Recurrent MDD	50.9	-	
Family history, recurrent MDD, or early age of onset (< 30 years)	94.8	-	

Table 2

Information on the SNPs with the smallest association p-values in the GWAS.

Basic SNP data					Bioinformatics			Results						MAF				Quality control - SNP missingness				Quality control -
SNP	Chr	Position	Alleles	Strand	Gene	TAMAL †	SLEP ‡	Rank	OR (CI)	P-asymptotic	P-empirical	q-value	P-gwemp	All	Cases	Controls	HapMap_EUR	All	Cases	Controls	p-missing	additional checks
rs12471796	2	20,177,820	A/G	+				10	1.26 (1.14-1.39)	0.000014	0.000014	0.58	0.99	0.298	0.322	0.275	0.271	0.012	0.010	0.014	0.36	
rs7565124	2	20,183,313	A/G	+		Reg pot		7	1.26 (1.14-1.40)	0.000012	0.000011	0.58	0.98	0.296	0.321	0.272	0.271	0.030	0.034	0.026	0.20	
rs3923028	2	29,597,247	T/C	−	ALK	CNV	CNV, mutated in colon CA	12	1.34 (1.17-1.54)	0.000024	0.000020	0.66	1.00	0.135	0.153	0.119	0.175	0.001	0.002	0.000	0.06	
rs12621441	2	201,794,446	A/G	+			Near CNV	13	1.31 (1.16-1.49)	0.000024	0.000027	0.66	1.00	0.166	0.185	0.147	0.150	0.008	0.008	0.009	0.86	
rs11132183	4	184,428,336	T/C	+			MDD linkage peak (8.6 mb)	16	0.75 (0.65-0.86)	0.000029	0.000035	0.66	1.00	0.133	0.116	0.150	0.117	0.001	0.001	0.002	0.63	
rs17074623	4	184,652,456	G/A	+			MDD linkage peak (8.3 mb)	23	0.75 (0.66-0.86)	0.000043	0.000040	0.66	1.00	0.137	0.120	0.154	0.076	0.003	0.005	0.002	0.26	
rs2094927	6	14,397,061	T/G	−			SCZ linkage meta-analsyis (2.5 mb)	20	0.82 (0.74-0.90)	0.000042	0.000046	0.66	1.00	0.417	0.393	0.441	0.475	0.001	0.001	0.002	1.00	
rs2274822	6	14,399,068	C/T	−			SCZ linkage meta-analsyis (2.5 mb)	6	0.79 (0.71-0.88)	0.000009	0.000007	0.58	0.96	0.268	0.245	0.291	0.283	0.003	0.002	0.003	1.00	
rs1558477	7	30,928,587	C/T	+			MDD linkage peak (3.0 mb)	1	1.27 (1.16-1.40)	0.000001	0.000002	0.28	0.37	0.430	0.460	0.401	0.442	0.003	0.003	0.004	0.77	HWD cases
rs7791986	7	30,930,719	G/C	+			MDD linkage peak (3.0 mb)	14	1.22 (1.12-1.35)	0.000026	0.000038	0.66	1.00	0.451	0.477	0.427	0.425	0.001	0.001	0.002	0.38	

Mol Psychiatry. Author manuscript; available in PMC 2009 October 01.

Basic SNP data					Bioinformatics			Results						MAF				Quality control - SNP missingness				Quality cont
SNP	Chr	Position	Alleles	Strand	Gene	TAMAL †	SLEP ‡	Rank	OR (CI)	P-asymptotic	P-empirical	q-value	P-gwemp	All	Cases	Controls	HapMap_EUR	All	Cases	Controls	p-missing	additiona checks
rs2715148	7	82,094,686	A/C	+	PCLO	Cons, reg pot	BIP GWAS rs2715148 (p=0.03)	2	0.79 (0.72-0.87)	0.000001	0.000003	0.28	0.42	0.482	0.452	0.510	0.525	0.002	0.002	0.002	0.72	
rs2522833	7	82,098,359	C/A	+	PCLO	Cons, reg pot, cSNP	BIP GWAS rs7781142 (p=0.03)	3	1.26 (1.15-1.39)	0.000002	0.000002	0.28	0.52	0.455	0.485	0.427	0.425	0.002	0.000	0.003	0.03	
rs2522840	7	82,123,066	G/T	+	PCLO	Cons, reg pot	BIP GWAS rs7799260 (p=0.04)	4	1.25 (1.14-1.38)	0.000004	0.000003	0.40	0.74	0.456	0.484	0.428	0.425	0.004	0.002	0.006	0.18	
rs2107828	7	82,200,320	A/T	+	PCLO	Reg pot		8	0.81 (0.74-0.89)	0.000013	0.000007	0.58	0.99	0.460	0.433	0.486	0.500	0.037	0.036	0.038	0.79	
rs1457266	8	24,825,757	A/G	−		Reg pot	MDD linkage peak (7.4 mb)	17	0.81 (0.73-0.89)	0.000029	0.000034	0.66	1.00	0.319	0.295	0.342	0.300	0.002	0.002	0.001	0.44	
rs7005189	8	81,663,211	T/C	+		Cons, reg pot	BIP GWAS rs11778905 (p=0.03, 9.9 kb)	15	0.76 (0.66-0.86)	0.000028	0.000036	0.66	1.00	0.153	0.134	0.170	0.150	0.001	0.000	0.002	0.25	
rs1780436	10	34,297,618	A/G	−				5	0.80 (0.73-0.88)	0.000008	0.000013	0.58	0.95	0.374	0.348	0.400	0.325	0.018	0.016	0.021	0.31	
rs11031676	11	32,242,721	T/C	+		Reg pot		21	1.26 (1.13-1.40)	0.000043	0.000035	0.66	1.00	0.232	0.253	0.212	0.139	0.006	0.007	0.004	0.28	
rs12579771	12	44,019,689	T/C	+	TMEM16F	Cons, reg pot	MDD linkage peak (8.8 mb)	11	0.78 (0.69-0.87)	0.000022	0.000023	0.66	1.00	0.205	0.184	0.225	0.271	0.000	0.000	0.001	1.00	
rs4765078	12	123,171,707	C/T	+				25	0.82 (0.74-0.90)	0.000044	0.000035	0.66	1.00	0.374	0.350	0.397	0.408	0.004	0.003	0.004	1.00	
rs8023445	15	46,980,083	C/T	+	SHC4	Reg pot		9	0.72 (0.62-0.84)	0.000014	0.000009	0.58	0.99	0.119	0.101	0.135	0.108	0.011	0.012	0.010	0.63	
rs3885179	19	14,688,830	A/C	−	ZNF333	Reg pot, cSNP		18	0.61 (0.48-0.77)	0.000032	0.000031	0.66	1.00	0.046	0.035	0.056	0.033	0.021	0.024	0.018	0.24	
rs941796	20	39,724,220	A/G	+		Reg pot		22	1.22 (1.11-1.35)	0.000043	0.000037	0.66	1.00	0.398	0.422	0.374	0.408	0.013	0.013	0.012	0.88	
rs12480143	20	39,741,240	G/A	+				24	1.25 (1.13-1.39)	0.000044	0.000035	0.66	1.00	0.265	0.288	0.244	0.233	0.001	0.001	0.002	1.00	

Mol Psychiatry. Author manuscript; available in PMC 2009 October 01.

Basic SNP data					Bioinformatics			Results						MAF				Quality control - SNP missingness				Quality cont	
SNP	Chr	Position	Alleles	Strand	Gene	TAMAL †	SLEP ‡	Rank	OR (CI)	P-asymptotic	P-empirical	q-value	P-gwemp	All	Cases	Controls	HapMap_EUR	All	Cases	Controls	p-missing	Sullivan et al.	additiona checks
rs928862	21	20,559,590	G/A	+		Reg pot	Near CNV	19	0.78 (0.69-0.88)	0.000040	0.000044	0.66	1.00	0.190	0.170	0.209	0.167	0.013	0.012	0.013	0.77		

Notes. Sorted by location. All locations per NCBI Build 35 (UCSC hg17). Alleles are given as minor/major. OR (CI)=odds ratio (95% confidence interval). P-asymptomic=p-value from Trend test. P-empirical=pointwise p-value from adaptive permuation method in PLINK. q-value=see text. P-gwemp=genomewide empirical p-value via traditional permutation testing (5000 replicates). MAF=minor allele frequency. HapMap MAFs have been converted to the reference allele of the MDD sample. P-missing tests the difference in missingness between cases and controls. For noteworthy associations, the four flags refer to acceptable cluster plots, conformation to Hardy-Weinberg equilibrium, absence of plate-specific association outliers, and the presence of a “proxy” SNP in high linkage disequilibrium with the primary SNP.

†TAMAL codes. cSNP=coding SNP. Bioinformatic flag possibilities: coding SNP (cSNP), SNP in segmental duplication, known copy number variant (CNV), conserved base (Cons), miRNA target site, region of regulatory potential (reg pot), predicted promoter, transfactor binding site, enhancer, exon, splice site, mRNA expression QTL (lymphocytes or cortex). Only positive flags are shown.

‡SLEP=Sullivan Lab Evidence Project (http://slep.unc.edu) a compendium of genetic findings from the literature. Sources (PubMed IDs): CNVs from Database of Genomic Variation (15286789), breast and colon cancer mutations (17932254), MDD genomewide linkage studies (12612864, 14582139, 17427203), SCZ genomewide linkage meta-analysis (12802786), and bipolar disorder (BIP) GWAS (17554300).

Table 3

Clustering of SNPs with low p-values.

Rank	Chr	Start	End	Nsnps	Pmin	N <0.0001	N <0.001	N <0.01	Expressed in brain?	Genes	Gene products	SLEP ‡
1	7	30,928,587	30,931,521	3	1.25E-06	2	0	1	Yes	ADCYAP1R1	adenylate cyclase activating polypeptide 1 (pituitary) receptor type I	Neuroactive ligand-receptor interaction
2	7	82,041,576	82,208,167	10	1.50E-06	6	4	0	Yes	PCLO	piccolo (presynaptic cytomatrix protein)	
4	6	14,388,932	14,399,068	2	9.09E-06	1	1	0				
5	2	20,177,820	20,183,313	2	1.18E-05	2	0	0	Yes	LAPTM4A	lysosomal-associated protein transmembrane 4 alpha	
6	15	46,979,618	46,980,083	2	1.36E-05	1	1	0	Yes	*CRI1 EID1 RaLP SHC4*	CREBBP/EP300 inhibitor 1/EP300 interacting inhibitor of differentiation 1/rai-like protein/SHC (Src homology 2 domain containing) family, member 4	
9	2	201,794,446	201,880,818	2	2.44E-05	1	1	0	Yes	*AJ487678 AJ487679 AK125394 AY690601 CASP10 CFLAR NDUFB3*	Caspase 10./Caspase 10./PRO3098./Caspase 10 splice variant G./caspase 10, apoptosis-related cysteine peptidase/CASP8 and FADD-like apoptosis regulator/NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 3, 12kDa	CASP10 causes multiple neoplasms (OMIM 601762); CFLAR upregulated in MDD in post-mortem brain
14	20	39,724,220	39,742,644	5	4.27E-05	5	0	0				
15	6	14,386,148	14,397,061	3	4.23E-05	1	1	1				
16	4	184,652,456	184,658,003	3	4.28E-05	1	0	2				
17	5	117,174,763	117,282,887	4	4.84E-05	1	1	2				
19	10	127,071,672	127,087,021	3	0.000046	1	2	0				
20	5	22,752,605	22,792,155	3	4.65E-05	1	0	2	Yes	CDH12	cadherin 12, type 2 (N-cadherin 2)	
22	15	88,130,196	88,136,792	2	5.52E-05	1	1	0	Yes	ANPEP MESP2	alanyl (membrane) aminopeptidase (aminopeptidase N, aminopeptidase M, microsomal aminopeptidase, CD13, p150)/mesoderm posterior 2 homolog (mouse)	MESP2 causes spondylocostal dysostosis (OMIM 605195)
23	8	54,098,247	54,102,064	2	4.83E-05	2	0	0				
24	4	145,875,183	145,878,794	2	5.47E-05	1	0	1				
27	11	32,242,721	32,244,520	2	4.25E-05	2	0	0				
28	8	27,249,840	27,379,524	6	5.38E-05	1	2	3	Yes	*AK128371 CHRNA2 PTK2B*	Hypothetical protein FLJ46514./cholinergic receptor, nicotinic, alpha 2 (neuronal)/PTK2B protein tyrosine kinase	CHRNA2 causes nocturnal frontal lobe epilepsy (OMIM 118502)

Rank	Chr	Start	End	Nsnps	Pmin	N <0.0001	N <0.001	N <0.01	Expressed in brain?	Genes	Gene products	SLEP ‡
											2 beta	
29	3	12,453,817	12,459,985	2	0.00005	1	1	0		PPARG	peroxisome proliferator-activated receptor gamma	Type 2 diabetes mellitus risk gene
32	3	99,975,821	100,183,009	2	7.26E-05	1	0	1	Yes	DCBLD2 ST3GAL6	discoidin, CUB and LCCL domain containing 2/ST3 beta-galactoside alpha-2,3-sialyltransferase 6	
34	3	70,451,852	70,476,913	2	8.22E-05	1	0	1				
38	2	7,424,098	7,440,754	3	8.95E-05	1	1	1				
41	5	54,352,635	54,363,712	3	0.000071	1	1	1	Yes	GZMK	granzyme K (granzyme 3; tryptase II)	
43	13	111,889,281	111,902,203	2	7.87E-05	1	1	0				
44	1	211,470,329	211,508,991	4	0.000072	1	3	0				
46	8	24,784,576	24,825,085	2	0.000359	0	2	0		NEF3 NEFM	neurofilament 3 (150kDa medium)/neurofilament, medium polypeptide 150kDa	

‡SLEP=Sullivan Lab Evidence Project (http://slep.unc.edu) a compendium of genetic findings from the literature.

Table 4

PCLO replication results.

(a) MDD (major depressive disorder).

SNP ID	hg17 position	NESDA-NTR (NL)			All replication samples					QIMR (AUS)				M-P Inst Psych (GER)				West Germany (GER)				STAR*D (US)				U Edinburgh (UK)				DeCC (UK)			
		N	Z	P	N	Z	P	P_corr	I^2	N	Z	P	P_corr	N	Z	P	P_corr	N	Z	P	P_corr	N	Z	P	P_corr	N	Z	P	P_corr	N	Z	P	P_corr
rs7780196	82,021,603	3,657	2.87	0.00406	9,270	1.58	0.11	0.81	0.65	2,004	2.87	0.0041	0.066	1,907	1.87	0.06	0.59	2,482	-0.82	0.41	1	2,051	-0.91	0.36	1	826	0.68	0.49	1				
rs17282875	82,041,576	3,657	-4.09	4.3E-05	7,235	-0.23	0.82	1	0.61	2,002	-2.31	0.021	0.29					2,479	0.55	0.58	1	1,928	1.40	0.16	0.90	826	0.25	0.81	1				
rs10954689	82,047,024	3,657	3.91	9.1E-05	9,269	1.38	0.17	0.92	0.52	2,004	2.88	0.0039	0.076	1,899	0.54	0.59	1	2,489	0.76	0.45	1	2,051	-1.10	0.27	0.98	826	-0.12	0.90	1				
rs12672552	82,058,350	3,658	0.58	0.56	9,277	-1.38	0.17	0.92	0.00	2,005	-0.30	0.77	1	1,907	-0.81	0.42	1	2,488	-0.02	0.99	1	2,051	-0.98	0.32	0.99	826	-1.37	0.17	0.91				
rs6948464	82,061,983	3,658	-3.46	0.00055	9,269	0.86	0.39	1	0.00	2,005	-1.05	0.29	0.99	1,907	0.46	0.65	1	2,480	0.56	0.58	1	2,051	1.50	0.13	0.85	826	0.92	0.36	1				
rs13227462	82,065,698	3,658	-0.98	0.32	9,271	-0.81	0.42	1	0.69	2,004	-0.24	0.81	1	1,906	-1.42	0.16	0.87	2,484	0.56	0.58	1	2,051	1.08	0.28	0.98	826	-2.88	0.0040	0.050				
rs17156675	82,067,056	3,658	3.89	0.0001	9,277	1.61	0.11	0.80	0.16	2,005	2.48	0.013	0.20	1,907	0.70	0.48	1	2,488	0.74	0.46	1	2,051	-0.21	0.84	1	826	-0.39	0.69	1				
rs6979066	82,067,588	3,656	3.94	8.1E-05	9,237	1.81	0.07	0.66	0.39	1,983	2.65	0.0082	0.13	1,906	0.93	0.35	1	2,475	0.91	0.36	1	2,047	-0.99	0.32	0.99	826	0.48	0.63	1				
rs6965452	82,073,522	3,658	-1.51	0.13	9,267	-1.34	0.18	0.92	0.00	2,001	-1.97	0.048	0.51	1,903	-0.88	0.38	1	2,486	-0.42	0.68	1	2,051	0.62	0.54	1	826	-0.74	0.46	1				
rs11771757	82,074,382	3,645	0.68	0.50	6,774	0.48	0.63	1	0.66	1,993	0.83	0.41	1	1,905	0.63	0.53	1					2,051	-1.85	0.06	0.59	825	1.93	0.05	0.55				
rs12668093	82,079,684	3,658	0.73	0.47	9,258	-2.21	0.027	0.36	0.00	2,005	-0.38	0.71	1	1,888	-0.99	0.32	0.99	2,489	-1.53	0.13	0.82	2,050	-1.25	0.21	0.95	826	-0.97	0.33	0.99				
rs6954078	82,080,529	3,633	-4.44	8.8E-06	9,204	-0.07	0.94	1	0.48	1,990	-2.26	0.024	0.33	1,890	-0.18	0.85	1	2,469	0.79	0.43	1	2,037	1.38	0.17	0.91	818	0.17	0.86	1				
rs2715148	82,094,686	3,651	-5.05	4.4E-07	11,850	0.23	0.82	1	0.40	2,003	-1.77	0.076	0.66	1,905	-0.66	0.51	1	2,478	0.68	0.49	1	2,048	2.06	0.04	0.45	826	0.16	0.88	1	2,590	0.33	0.74	0.88
rs2522833	82,098,359	3,658	5.01	5.4E-07	11,934	0.12	0.90	1	0.41	2,002	2.20	0.028	0.35	1,892	0.49	0.62	1	2,484	0.35	0.72	1	2,050	-1.44	0.15	0.89	826	-0.10	0.92	1	2,680	-1.10	0.27	0.37
rs2522840	82,123,066	3,656	4.79	1.6E-06	9,268	0.63	0.53	1	0.40	2,004	2.06	0.039	0.44	1,904	0.55	0.58	1	2,483	0.33	0.75	1	2,051	-1.52	0.13	0.84	826	-0.16	0.87	1				
rs13233504	82,142,482	3,656	2.63	0.00843	9,263	0.82	0.41	1	0.12	2,003	1.12	0.26	0.98	1,907	1.49	0.14	0.84	2,476	-0.22	0.83	1.00	2,051	-1.08	0.28	0.99	826	0.56	0.58	1				
rs2888018	82,145,941	3,657	3.80	0.00015	9,260	-0.86	0.39	1	0.19	2,005	0.60	0.55	1	1,903	0.73	0.47	1	2,477	-1.23	0.22	0.95	2,049	-1.85	0.07	0.62	826	-0.40	0.69	1				
rs2371364	82,151,525	3,656	2.22	0.02611	9,271	1.60	0.11	0.80	0.21	2,004	2.64	0.008	0.13	1,903	-0.33	0.74	1	2,487	0.69	0.49	1	2,051	0.09	0.93	1	826	0.24	0.81	1				
rs2371367	82,163,042	3,656	2.87	0.00406	9,257	0.67	0.51	1	0.00	2,004	0.88	0.38	1	1,893	-0.32	0.75	1	2,484	0.13	0.89	1	2,050	0.08	0.94	1	826	0.08	0.94	1				
rs2189972	82,169,314	3,657	1.37	0.17	9,272	2.20	0.028	0.35	0.00	2,004	1.69	0.092	0.71	1,905	0.11	0.91	1	2,486	0.85	0.40	1	2,051	0.80	0.42	1	826	1.16	0.25	0.98				
rs17235252	82,180,688	3,658	-1.03	0.30	7,363	-1.70	0.09	0.73	0.62	2,004	-1.47	0.14	0.87					2,487	-2.29	0.02	0.26	2,046	-0.05	0.96	1	826	1.47	0.14	0.86				
rs17809157	82,192,478	3,652	-2.58	0.00975	9,266	-1.94	0.05	0.56	0.00	2,003	-1.36	0.17	0.92	1,906	-1.53	0.13	0.82	2,481	-1.15	0.25	0.97	2,050	0.26	0.79	1	826	0.47	0.64	1				
rs2107828	82,200,320	3,658	-4.69	2.8E-06	9,271	-1.05	0.29	0.99	0.09	2,005	-1.59	0.11	0.79	1,902	-1.09	0.27	0.98	2,487	-0.30	0.76	1	2,051	0.97	0.33	0.99	826	0.48	0.63	1				
rs10954694	82,201,812	3,656	2.77	0.00558	9,274	2.53	0.011	0.19	0.00	2,005	1.24	0.22	0.95	1,907	1.02	0.31	0.99	2,486	1.59	0.11	0.77	2,050	0.71	0.47	1	826	0.75	0.45	1				
rs10487645	82,203,942	3,658	0.44	0.66	9,272	-0.79	0.43	1	0.00	2,005	0.17	0.87	1	1,901	0.55	0.58	1	2,489	-1.87	0.06	0.55	2,051	-0.71	0.48	1	826	-0.15	0.88	1				
rs9690648	82,205,975	3,658	0.52	0.60	9,278	-1.01	0.31	0.99	0.01	2,005	-0.01	0.99	1	1907.00	0.69	0.49	1	2,489	-2.02	0.04	0.43	2,051	-0.97	0.33	0.99	826	-0.24	0.81	1				
rs17235831	82,206,612	3,652	-0.94	0.35	9,271	-0.75	0.46	1	0.22	2,003	-0.69	0.49	1	1,904	-0.43	0.67	1	2,488	-1.34	0.18	0.91	2,051	0.16	0.88	1	825	1.57	0.12	0.80				
rs6959723	82,206,991	3,656	-2.16	0.03084	9,265	-1.94	0.05	0.55	0.00	2,003	-0.70	0.48	1	1,905	-1.30	0.19	0.93	2,481	-1.15	0.25	1	2,050	-0.42	0.67	1	826	-0.44	0.66	1				
rs7799260	82,208,167	3,658	-4.19	2.8E-05	9,274	-0.87	0.39	1	0.00	2,005	-0.78	0.44	1	1,905	-0.94	0.35	0.99	2,487	-0.30	0.76	1	2,051	0.51	0.61	1	826	0.41	0.68	1				
rs12669254	82,217,749	3,656	2.31	0.02077	9,274	-2.06	0.039	0.46	0.00	2,005	-0.93	0.35	0.99	1,904	-0.58	0.56	1	2,489	-0.82	0.41	1	2,050	-1.04	0.30	0.99	826	-1.79	0.07	0.62				

(b) reoMDD (recurrent, early-onset MDD).

SNP ID	hg17 position	NESDA-NTR (NL)			All replication samples					QIMR (AUS)				M-P Inst Psych (GER)				West Germany (GER)				STAR*D (US)				U Edinburgh (UK)				DeCC (UK)			
		N	Z	P	N	Z	P	P_corr	I^2	N	Z	P	P_corr	N	Z	P	P_corr	N	Z	P	P_corr	N	Z	P	P_corr	N	Z	P	P_corr	N	Z	P	P_corr
rs7780196	82,021,603	2,369	2.80	0.00514	6,403	0.86	0.39	1	0.64	1,255	2.16	0.03	0.36	1,310	2.05	0.04	0.45	1,580	-1.16	0.24	0.97	1,548	-1.09	0.28	0.98	710	0.17	0.86	1				
rs17282875	82,041,576	2,368	-4.48	7.5E-06	4,983	0.10	0.92	1	0.00	1,254	-1.00	0.32	0.99					1,582	0.68	0.50	1	1,437	1.06	0.29	0.99	710	0.20	0.84	1				
rs10954689	82,047,024	2,369	4.06	4.9E-05	6,403	1.22	0.22	0.97	0.00	1,255	1.27	0.20	0.95	1,305	1.09	0.28	0.98	1,585	0.11	0.91	1	1,548	-0.31	0.76	1	710	0.25	0.81	1				
rs12672552	82,058,350	2,369	1.19	0.24	6,409	-0.50	0.61	1	0.00	1,256	-1.28	0.20	0.94	1,310	-0.30	0.76	1	1,585	0.22	0.83	1	1,548	0.18	0.86	1	710	-1.08	0.28	0.99				
rs6948464	82,061,983	2,369	-2.86	0.00430	6,407	0.93	0.35	1	0.00	1,256	-0.58	0.56	1	1,310	-0.05	0.96	1	1,583	1.76	0.08	0.67	1,548	1.16	0.25	0.97	710	0.86	0.39	1				
rs13227462	82,065,698	2,369	-2.96	0.00306	6,405	-0.91	0.36	1	0.54	1,255	-0.23	0.82	1	1,309	-1.93	0.05	0.51	1,583	0.15	0.88	1	1,548	0.90	0.37	1	710	-2.21	0.027	0.296				
rs17156675	82,067,056	2,369	3.88	1.1E-04	6,409	1.15	0.25	0.98	0.27	1,256	2.40	0.02	0.25	1,310	0.70	0.49	1	1,585	0.47	0.64	1	1,548	-0.41	0.68	1	710	-0.23	0.82	1				
rs6979066	82,067,588	2,367	3.64	2.7E-04	6,383	1.13	0.26	0.98	0.44	1,238	2.09	0.04	0.43	1,309	1.37	0.17	0.91	1,580	0.24	0.81	1	1,546	-1.03	0.30	0.99	710	0.58	0.56	1				
rs6965452	82,073,522	2,369	-3.16	0.00155	6,401	-1.14	0.26	0.98	0.00	1,253	-1.15	0.25	0.97	1,307	-1.12	0.26	0.97	1,583	-1.04	0.30	0.98	1,548	0.43	0.67	1	710	-0.69	0.49	1				
rs11771757	82,074,382	2,361	-0.02	0.99	4,814	0.84	0.40	1	0.60	1,249	-0.33	0.74	1	1,308	1.58	0.11	1					1,548	-1.39	0.16	0.90	709	1.70	0.09	0.72				
rs12668093	82,079,684	2,369	1.92	0.06	6,395	-1.64	0.10	0.77	0.00	1,256	-0.97	0.33	0.99	1,296	-0.68	0.50	1	1,585	-0.94	0.35	0.99	1,548	-0.64	0.52	1	710	-1.28	0.20	0.94				
rs6954078	82,080,529	2,355	-4.98	6.5E-07	6,365	0.10	0.92	1	0.21	1,251	-1.28	0.20	0.95	1,300	-0.78	0.44	1	1,578	1.35	0.18	0.91	1,534	1.01	0.31	0.99	702	0.30	0.76	1				
rs2715148	82,094,686	2,363	-5.35	9.0E-08	8,416	0.32	0.75	1	0.40	1,255	-0.85	0.40	1	1,308	-1.17	0.24	0.96	1,577	1.77	0.08	0.68	1,547	1.92	0.06	0.55	710	0.19	0.85	1	2,019	-0.18	0.86	0.96
rs2522833	82,098,359	2,369	4.95	7.4E-07	8,478	0.21	0.83	1	0.03	1,254	1.09	0.28	0.98	1,300	1.39	0.16	0.88	1,582	-0.76	0.45	1	1,547	-0.99	0.32	0.99	710	0.14	0.89	1	2,085	-0.48	0.63	0.78
rs2522840	82,123,066	2,368	4.89	9.9E-07	6,404	0.46	0.65	1	0.18	1,256	1.12	0.26	0.98	1,308	1.32	0.19	0.91	1,582	-0.74	0.46	1	1,548	-1.08	0.28	0.98	710	0.14	0.89	1				
rs13233504	82,142,482	2,367	2.26	0.02405	9,405	0.50	0.62	1	0.08	1,256	-0.05	0.96	1	1,310	1.73	0.08	0.67	1,581	-0.65	0.51	1	1,548	-0.84	0.40	1	710	0.44	0.66	1				
rs2888018	82,145,941	2,368	4.08	4.5E-05	6,399	-1.26	0.21	0.95	0.15	1,256	-1.06	0.29	0.98	1,306	1.11	0.27	0.97	1,579	-1.55	0.12	0.81	1,548	-1.54	0.12	0.82	710	-0.62	0.54	1				
rs2371364	82,151,525	2,368	1.43	0.15	6,403	1.15	0.25	0.98	0.58	1,256	3.23	0.001	0.02	1,306	-0.68	0.49	1	1,583	-0.29	0.78	1	1,548	0.18	0.85	1	710	0.47	0.64	1				
rs2371367	82,163,042	2,367	2.40	0.02	6,397	0.14	0.89	1	0.00	1,256	0.10	0.92	1	1,303	-0.18	0.86	1	1,581	-1.21	0.23	0.96	1,547	-0.07	0.95	1	710	0.35	0.73	1				
rs2189972	82,169,314	2,369	0.94	0.35	6,406	1.44	0.15	0.89	0.00	1,255	1.67	0.10	0.73	1,310	-0.08	0.94	1	1,583	-0.20	0.84	1	1,548	0.21	0.83	1	710	1.00	0.32	0.99				
rs17235252	82,180,688	2,369	-1.27	0.20	5,096	-0.92	0.36	1	0.36	1,256	-0.50	0.62	1					1,585	-1.78	0.08	0.62	1,545	-0.02	0.99	1	710	1.25	0.21	0.95				
rs17809157	82,192,478	2,364	-1.96	0.05	6,402	-0.49	0.62	1	0.00	1,255	-0.68	0.49	1	1,309	-0.68	0.49	1	1,581	0.86	0.39	1	1,547	0.54	0.59	1	710	0.59	0.55	1				
rs2107828	82,200,320	2,369	-4.60	4.2E-06	8,405	-0.06	0.95	1	0.00	1,256	-0.56	0.58	1	1,307	-1.04	0.30	0.98	1,584	1.14	0.25	0.97	1,548	1.18	0.24	0.97	710	0.38	0.71	1				
rs10954694	82,201,812	2,367	1.91	0.06	6,406	2.51	0.01	0.19	0.00	1,256	1.72	0.09	0.71	1,310	1.13	0.26	0.97	1,583	0.70	0.48	1	1,547	0.48	0.63	1	710	0.92	0.36	1				
rs10487645	82,203,942	2,369	1.55	0.12	6,405	-1.49	0.14	0.86	0.00	1,256	-1.12	0.26	0.97	1,306	0.41	0.68	1	1,585	-1.01	0.31	0.99	1,548	-1.16	0.25	0.97	710	-0.94	0.34	1				
rs9690648	82,205,975	2,369	1.74	0.08	6,409	-1.68	0.09	0.73	0.00	1,256	-1.22	0.22	0.95	1,310	0.43	0.67	1	1,585	-1.13	0.26	0.97	1,548	-1.47	0.14	0.86	710	-0.85	0.39	1				
rs17235831	82,206,612	2,364	-1.03	0.30	6,403	-0.25	0.80	1	0.00	1,255	0.17	0.87	1	1,307	0.36	0.72	1	1,584	-1.43	0.15	0.87	1,548	-0.04	0.97	1	709	1.21	0.23	0.96				
rs6959723	82,206,991	2,368	-3.11	0.00189	6,402	-1.23	0.22	0.96	0.00	1,255	-0.02	0.98	1	1,308	-1.08	0.28	0.98	1,582	-1.38	0.17	0.90	1,547	-0.12	0.90	1	710	-0.49	0.63	1				
rs7799260	82,208,167	2,369	-4.30	1.7E-05	6,407	-0.61	0.54	1	0.00	1,256	0.19	0.85	1	1,309	-1.20	0.23	0.95	1,584	0.33	0.74	1	1,548	0.56	0.58	1	710	0.35	0.73	1				
rs12669254	82,217,749	2,368	2.50	0.01	6,406	-1.83	0.07	0.63	0.04	1,256	-2.17	0.03	0.35	1,308	-0.15	0.88	1	1,585	-0.98	0.33	0.99	1,547	-0.40	0.69	1	710	-1.23	0.22	0.95				

Abbreviations: N=total sample size for an analysis, Z=logistic regression beta divided by its standard error, P=asymptotic p-value from Wald chi-square test (1 df) uncorrected for multiple comparisons, P_corr=empirical p-value from accounting for multiple comparisons and LD structure (50K permutations), and I^2=i-squared, an index of heterogeneity of logistic regression parameter estimates.