Published in final edited form as:

Bipolar Disord. 2014 September ; 16(6): 608–616. doi:10.1111/bdi.12201.

Bipolar polygenic loading and bipolar spectrum features in major depressive disorder

Anna Wiste^{a,b}, Elise B Robinson^c, Yuri Milaneschi^d, Sandra Meier^e, Stephan Ripke^c, Caitlin C Clements^{a,b}, Garrett M Fitzmaurice^f, Marcella Rietschel^e, Brenda W Penninx^{d,g,h}, Jordan W Smoller^{a,i}, and Roy H Perlis^{a,b}

^aPsychiatric and Neurodevelopmental Genetics Unit, Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA ^bCenter for Experimental Drugs and Diagnostics, Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA

^cAnalytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA

^dDepartment of Psychiatry, VU University Medical Center, Amsterdam, The Netherlands

^eDepartment of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical

Faculty Mannheim/Heidelberg University, Mannheim, Germany ^fDepartment of Biostatistics,

Harvard School of Public Health, Boston, MA, USA ^gDepartment of Psychiatry, Leiden University

Medical Center, Leiden ^hDepartment of Psychiatry, University Medical Center of Groningen, Groningen, The Netherlands ⁱDepartment of Psychiatry, Massachusetts General Hospital, Boston, MA, USA

Abstract

Objectives—Family and genetic studies indicate overlapping liability for major depressive disorder and bipolar disorder. The purpose of this study was to determine whether this shared genetic liability influences clinical presentation.

Methods—A polygenic risk score for bipolar disorder, derived from a large genome-wide association meta-analysis, was generated for each subject of European–American ancestry ($n = 1,274$) in the Sequential Treatment Alternatives to Relieve Depression study (STAR*D) outpatient major depressive disorder cohort. A hypothesis-driven approach was used to test for association between bipolar disorder risk score and features of depression associated with bipolar disorder in the literature. Follow-up analyses were performed in two additional cohorts.

Results—A generalized linear mixed model including seven features hypothesized to be associated with bipolar spectrum illness was significantly associated with bipolar polygenic risk score [$F = 2.07$, degrees of freedom (df) = 7, $p = 0.04$]. Features included early onset, suicide attempt, recurrent depression, atypical depression, subclinical mania, subclinical psychosis, and

Corresponding author: Roy H. Perlis, MD, MSc, Psychiatric and Neurodevelopmental Genetics Unit, Center for Human Genetic Research, Simches Research Building, Massachusetts General Hospital, 185 Cambridge Street, Boston, MA 02114, USA, Fax: 617-726-0830, rperlis@chgr.mgh.harvard.edu.

Disclosures

JWS has served as a consultant for The Medical Letter. RHP is a member of scientific advisory boards or has received consulting fees from Genomind, Healthraceous, PamLab, Proteus Biomedical, and RIDVentures; and has received research support from Proteus Biomedical, and royalties from Concordant Rater Systems (now UBC). AW, EBR, YM, SM, SR, CCC, GMF, MR, and BWP have no conflicts of interest to report.

severity. Post-hoc univariate analyses demonstrated that the major contributors to this omnibus association were onset of illness at age 18 years [odds ratio (OR) = 1.2, $p = 0.003$], history of suicide attempt (OR = 1.21, $p = 0.03$), and presence of at least one manic symptom (OR = 1.16, $p = 0.02$). The maximal variance in these traits explained by polygenic score ranged from 0.8–1.1%. However, analyses in two replication cohorts testing a five feature model did not support this association.

Conclusions—Bipolar genetic loading appeared to be associated with bipolar-like presentation in major depressive disorder in the primary analysis. However, results are at most inconclusive because of lack of replication. Replication efforts are challenged by different ascertainment and assessment strategies in the different cohorts. The methodological approach described here may prove useful in applying genetic data to clarify psychiatric nosology in future studies.

Keywords

bipolar disorder; depression; genetics; major depressive disorder; polygenic score; STAR*D

Understanding the relationship between bipolar disorder and major depressive disorder remains a nosologic challenge with substantial clinical implications. Indeed, Kraepelin (1) considered bipolar disorder and major depressive disorder to be aspects of a common underlying illness. More recently, a bipolar spectrum disorder has been proposed to capture those individuals with depression who have risk factors and symptomatic presentation similar to bipolar depression (2). Features of depression that may be associated with bipolarity include family history of bipolar disorder (3, 4), an earlier age of onset of depression (3, 4), suicidality (3, 5), more severe episodes occurring with greater frequency (3), irritability (6, 7), and atypical features (8), particularly neurovegetative symptoms. Another feature associated with bipolarity in the literature is cyclothymia (9, 10). A history of subclinical manic or psychotic (8, 11) symptoms in major depressive disorder may also be suggestive of risk factors shared with bipolar disorder. The presence of some of these bipolar-like features in major depressive disorder has been associated with poorer treatment outcomes (12).

Recent publications have identified multiple genome-wide significant genetic variants that are risk factors for bipolar disorder (13) and schizophrenia (14). Furthermore, these studies have demonstrated polygenicity of bipolar disorder and schizophrenia showing that polygenic scores capture statistically significant proportions of the risk for these disorders (13–15). By demonstrating that polygenic risk score for schizophrenia predicted risk of bipolar disorder, Purcell et al. (15) showed that polygenic scores can be used to predict risk for disorders with shared genetic risk. Polygenic scores provide an opportunity to directly examine the hypothesis that bipolar genetic loading—defined not by family history, but by burden of common risk variants identified in population studies—is associated with clinical features representative of bipolar spectrum illness among individuals with major depressive disorder.

The Sequential Treatment Alternatives to Relieve Depression study (STAR*D) cohort (16) includes patients treated for depression in outpatient settings. We predicted that in subjects with major depressive disorder, polygenic scores indicating higher genetic loading for

bipolar disorder would be associated with higher likelihood of exhibiting those features of depression associated with bipolar disorder. Two cohorts were identified for replication efforts. The Netherlands Study of Depression and Anxiety (NESDA) and Mannheim cohorts are also clinical studies of major depressive disorder for which genome-wide data and various bipolar risk phenotypes are available.

Methods

Subjects

Subjects in the primary analysis participated in the STAR*D treatment study of depression. Details of the cohort have previously been described extensively (16). All subjects were diagnosed with major depressive disorder by clinical interview and by DSM-IV symptoms. All subjects had a baseline score of ≥ 14 on the Hamilton Rating Scale for Depression (HRSD). DNA was collected from a subset of participants, and this analysis includes only those of European-American ancestry ($n = 1,274$). Mean age was 43.4 ± 13.5 years, and 59% were female.

Relevant exclusionary criteria for the study included lifetime diagnosis of psychotic depression, schizophrenia, schizoaffective disorder, bipolar I disorder, bipolar II disorder, or bipolar disorder not otherwise specified (NOS). Exclusion was based on clinical assessment or self-report.

Replication cohorts

The NESDA (17) cohort includes cases with major depressive disorder ascertained from both primary care and outpatient specialist care facilities. Major depressive disorder was diagnosed using the Composite Interview Diagnostic Instrument (CIDI). Subjects with a primary diagnosis of a psychotic disorder, obsessive compulsive disorder, bipolar disorder, or severe substance use dependence were excluded. All subjects in this analysis are of European ancestry and have a current major depressive disorder diagnosis ($n = 992$). Mean age is 40.6 ± 12.1 . The cohort is 67% female.

The Mannheim cohort consists of consecutive admissions for inpatient treatment of major depressive disorder to the Central Institute of Mental Health in Mannheim, Germany (18). Diagnosis of major depressive disorder was made by consensus best-estimate procedure based on all available records including structured interview [Structured Clinical Interview for DSM-IV (SCID-I)], medical records, and family history. Subjects with bipolar disorder or schizophrenia were excluded. All subjects are of European ancestry ($n = 585$). Mean age is 47.5 ± 13.7 years, and 64% of subjects are female.

Genotyping and polygenic risk score generation

Genotyping methods have been described previously for the STAR*D (19), NESDA (20), and Mannheim cohorts (18). Briefly, all subjects in STAR*D were genotyped on Affymetrix arrays. Approximately half of the sample was genotyped at Affymetrix using the 500K array, while the other half was genotyped using the Affymetrix Array 5.0. A panel of single nucleotide polymorphisms (SNPs) present on both arrays was generated and used as a basis

for imputation. The NESDA cohort was genotyped using the Perlegen 600K array, while the Mannheim cohort used the Illumina 610K array. Additional genotypes in all cohorts were imputed using BEAGLE 3.3 (21), with European ancestry samples from Hapmap3 as the reference panel. Population stratification was addressed using principal components generated with EIGENSTRAT (22) and used as covariates in all analyses.

Polygenic risk scores for bipolar disorder were generated using the results of the recent meta-analysis from the Psychiatric Genome-Wide Association Study (GWAS) Consortium (PGC) (13). Several filters were used to generate the list of SNPs used to derive each score. First, the results from the bipolar disorder analysis, which used HapMap2 as a reference for imputation, were trimmed to include only those SNPs present in HapMap3. Additional refinement of SNPs included in the scoring follows steps used by Ripke et al. (14). SNPs with imputation quality scores < 0.9 or with minor allele frequency $< 2\%$ in controls were removed, as were all SNPs within the MHC region (Chr6: 25–35 MB). The remaining SNPs were then further trimmed using PLINK (23). Linkage disequilibrium (LD) information was used to weight the included SNPs toward those with the strongest signal in each LD group. An r^2 threshold of 0.25 was used, with a window of 500 kb. The resulting number of SNPs included as the basis for scoring was 104,186. Scores were generated using the log odds ratio (OR) to weight individual SNPs at eight thresholds based on the p-values in the discovery GWAS dataset: p-value thresholds (P_T) = 0.001, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5. Scores were calculated over a wide range of thresholds because different thresholds are most predictive depending on the genetic architecture of the disorder or trait studied. Polygenic risk scores were standardized for all analyses.

Predicted features

The phenotypic features predicted to be associated with bipolar disorder were gathered from the literature and included severity, age at onset, suicide attempts, psychotic or manic features, atypical depressive features, and recurrent episodes. In STAR*D these features were defined using the following measures. Baseline HRSD score was dichotomized with a score of 25 or greater used to define severe depression. Age at onset was determined retrospectively at clinical interview (24). Onset before age 18 was defined as early onset. This age range encompasses the childhood- and adolescent-onset groups previously defined for analysis in this cohort (24). Recurrent depression was defined as three or more episodes. Data for episode number was indeterminate for 180 individuals, which in STAR*D could suggest chronic fluctuation or episodes too numerous to quantify. By convention and to maximize sample size, these individuals were coded as recurrent for the primary analysis. Presence of psychotic or manic symptoms was available using data from the Psychiatric Diagnostic Screening Questionnaire (PDSQ) (25). Atypical depression was defined using the Inventory of Depressive Symptomatology (IDS) (26). The concept of bipolar spectrum disorder is a complex one, and not all features were assessed. Notably not present in this analysis were irritability and cyclothymic temperament.

For replication cohorts, we sought to harmonize clinical features to maximize the comparability of the model between different cohorts. However, each study includes different assessment tools. History of suicide attempt, onset at 18 or younger, and

recurrence, defined as three or more episodes, could be generated in all cohorts. The NESDA cohort included the IDS (26) and severity was defined using a cut-off of 48, which is comparable to a score of 25 or greater in the HRSD (27). In the Bonn/Mannheim cohort, severity was defined according to the Operational Criteria Checklist (OPCRIT) (28) of substantial decrease from premorbid functioning. Atypical depression was defined using the OPCRIT in the Bonn/Mannheim study and specific IDS items according to Novick et al. (29) in NESDA. For the assessment of manic symptoms, NESDA included the Mood Disorders Questionnaire (MDQ) (30), and a positive score was defined according to Hirschfeld et al. (31) (MDQ positive: score ≥ 7 , symptoms clustering in the same time period, and causing at least moderate problems). In the Bonn/Mannheim cohort, association with at least one OPCRIT symptom of mania was assessed in the examination of subclinical mania.

Statistical analyses

A challenge in investigations of multiple phenotypes is the need to control type I error inflation. We therefore utilized an approach in which we first tested a single multivariate model to determine whether there was an overall association between bipolar polygenic risk score and phenotypic variation in major depression that has previously been associated with bipolar disorder. If this initial test was nominally significant, we then planned to test the individual components of this model post-hoc to determine the extent to which each contributed to the association.

To maximize power in the multivariate model, we required that each included phenotypic measure be nominally correlated with at least one other phenotypic measure. Tetrachoric correlations were calculated for all predicted features of depression except individual symptoms (Table 1). A threshold tetrachoric correlation of 0.2 with at least one other outcome was required for inclusion. The first model included early onset, recurrent depression, severity, history of suicide attempt, psychotic symptoms, manic symptoms, and atypical depression. Age and sex were included as covariates in all analyses. To control for effects related to different sites of recruitment, the sites were divided into primary care clinics and tertiary care centers, and this variable was entered into all association analyses. This variable does not account for all possible variation secondary to recruitment sites. However, it accounts for the most significant difference between sites. Additionally, the first five components from the principal component analysis (PCA) were included to control for population stratification.

A multivariate analysis was used to examine whether there was an overall association of risk score for bipolar disorder with the set of predicted features indexing phenotypic variation in major depression. Generalized linear mixed models (GLMM) were employed to estimate the strength of the overall relationship. GLMM allow for an overall or omnibus test of association with multiple, correlated variables. GLMM was chosen for this case because the features of interest are categorical variables. GLMMs include a random effect to account for the within-individual correlation in the set of predicted measures. An omnibus F -statistic, with seven degrees of freedom (df), provides an overall test of association. A corresponding p -value of less than 0.05 indicates that the risk score is nominally associated with the set of

predicted variables. The model allows for individuals with missing data on some of the phenotypic variables to be included. The association between risk score and the set of predicted variables was tested at all eight discovery P_T . The analysis was conducted using SAS version 9.3 (32, 33).

Post-hoc analyses were performed with logistic regression using STATA 10.1 (34). To reduce multiple testing, we first examined the P_T which maximized association in the multivariate model. Next, because the strongest association P_T varies according to the genetic architecture of the trait/disorder studied (35), all thresholds were tested for those outcomes which contributed to the model.

Results from the multivariate model confirmed our concern that different patterns might be present for the variance explained at different P_T for different outcomes. The threshold $P_T = 0.1$ was chosen for additional analyses based on the results of the multivariate model, as this threshold was able to capture a portion of the effect for each contributing variable despite different underlying patterns. In order to ensure that a stronger effect at a different P_T was not missed, $p < 0.2$ at $P_T = 0.1$ in post-hoc analyses was used as a cutoff for pursuing analyses at other thresholds.

Replication efforts

As described previously, although attempts were made to harmonize phenotypic data wherever possible, assessments necessarily varied between the three cohorts. Some features were not available in all cohorts and therefore were dropped. Specifically, there was no common assessment of subthreshold manic or hypomanic symptoms across all three studies; the tools in each study are not only substantially different in content but also in their purpose. Because of the *a priori* significance of this feature, we decided to exclude it from the multivariate model and test each cohort's definition independently. The model carried into replication efforts therefore consisted of five features: early onset, recurrent depression, severity, atypical depression, and history of suicide attempt.

If this model was significant at any of the eight thresholds, post-hoc analyses for all features would be performed at that threshold. Any significant results would be further explored across all thresholds. All multivariate analyses were performed using SAS, as were univariate analyses in NESDA. The univariate analyses in the Mannheim cohort were performed using STATA.

Results

Multivariate model

Polygenic risk score for bipolar disorder predicted a multivariate model including known risk factors for bipolar disorder and features associated with bipolar depression in the STAR*D cohort of patients with major depression at $P_T = 0.1$ ($F = 2.07$, $df = 7$, $p = 0.04$). Results were similar across the eight P_T thresholds with more modest evidence of association seen at very low or very high thresholds.

We then examined each of the seven phenotypic features included in this model individually for association with polygenic score. Post-hoc univariate analyses, using logistic regression with variance explained expressed as Nagelkerke's pseudo R^2 , demonstrated association with $p < 0.05$ at that threshold for early onset (OR = 1.2, $p = 0.003$, pseudo $R^2 = 0.008$), history of suicide attempt (OR = 1.21, $p = 0.03$, pseudo $R^2 = 0.007$), and presence of at least one manic symptom (OR = 1.16, $p = 0.02$, pseudo $R^2 = 0.006$; see Table 2). At this threshold, recurrent depression also met the criteria we had set for examination at all values of P_T ($p < 0.2$).

For these four features, the entire range of P_T was then examined to assess whether the underlying risk architecture differed by phenotype. The pattern of effect sizes across a range of thresholds can be instructive as to the underlying architecture of the genetic influences on a phenotype. We observed distinct patterns for the different contributing features (see Fig. 1). All four features showed strongest association with risk score at other thresholds. The peak association for history of suicide attempt (OR = 1.26, $p = 0.006$, pseudo $R^2 = 0.01$) and presence of manic symptoms (OR = 1.18, $p = 0.006$, pseudo $R^2 = 0.008$) was at $P_T = 0.01$, while peak association for recurrent (OR = 1.14, $p = 0.04$, pseudo $R^2 = 0.005$) and early-onset depression (OR = 1.21, $p = 0.002$, pseudo $R^2 = 0.01$) was at $P_T = 0.3$ and $P_T = 0.2$, respectively. Early-onset and recurrent depression were the two most strongly correlated outcomes, with a tetrachoric correlation of 0.66 in this sample. The weaker effect seen for recurrent depression could be entirely due to the association with early onset. After controlling for early onset there is no effect of polygenic score on recurrence at $P_T = 0.3$ ($p = 0.4$). There is no association of polygenic score with severe depression, atypical depression, or psychotic symptoms.

Replication

Multivariate model—Because only five of the seven features of the original model were to be included for replication analyses, the model was refit in STAR*D to include those five features. The five-feature multivariate model was significant at $p < 0.05$ at seven of eight thresholds tested in the STAR*D cohort. Again the strongest result was at $P_T = 0.1$ ($F = 2.64$, $df = 5$, $p = 0.03$). The model was not significant at any threshold tested in the Mannheim cohort. In NESDA the model reached significance at two thresholds, $P_T = 0.01$ and 0.05 (Table 3).

Post-hoc analyses in NESDA revealed that the effect was driven entirely by a *protective* effect of the bipolar polygenic score on history of suicide attempt at $P_T = 0.01$ (OR = 0.69, $p = 0.002$, pseudo $R^2 = 0.023$). This result was consistent with a $p < 0.05$ for a protective effect at seven of eight thresholds tested. These results are in direct contrast to our results in STAR*D, where polygenic score was associated with *risk* for history of suicide attempt across all thresholds with a peak effect seen at $P_T = 0.01$ (OR = 1.26, $p = 0.006$, pseudo $R^2 = 0.01$).

Subclinical mania

Results for association analysis with subclinical mania are shown in Table 4. Because three different measures, capturing subtly different phenotypes, were used to assess mania, the

results are not directly comparable. In both replication cohorts, the polygenic score did predict higher likelihood of presence of manic features; however, the effect reached $p < 0.05$ at only one threshold in NESDA and at no thresholds in the Mannheim cohort. Similar or greater effect sizes were present in each of the replication cohorts (Table 4). However, statistical power is lower in each of these cohorts because of smaller sample size.

Finally, there was no association between history of bipolar disorder in a first-degree relative as reported by the subject and the bipolar polygenic score at any threshold in either STAR*D or Mannheim; data was not available in NESDA.

Discussion

In the STAR*D cohort, our results suggest a link between genetic loading for bipolar disorder and features of depression which have been previously suggested to be associated with bipolar disorder, with post-hoc analyses suggesting that early onset of depression, history of suicide attempt, and subclinical mania were the strongest drivers of that association. However, these results did not replicate in two additional large cohorts of individuals with major depression, one from the Netherlands, the other Germany.

Independent of these results, we have described a novel approach to examining phenotypic implications of polygenic loading across disorders, one of which considers multiple variables and therefore offers advantages over traditional univariate approaches. The multivariate approach allows the testing of several correlated features, minimizing type 1 error by reducing the number of tests necessary.

We hypothesized that subclinical manic features should be predicted by bipolar polygenic loading if what we are assessing is truly a subclinical form of mania, rather than perhaps normal variation unrelated to true mania. In all three cohorts, which used distinctly different definitions of subclinical mania, dependent on the tools used to assess it, the polygenic score did tend to predict presence of manic features. However, the replication cohorts were both smaller than the discovery cohort with concomitant decrease in statistical power. Even though in each of these cohorts there were thresholds where the effect sizes were larger than any seen in the STAR*D cohort, there was little evidence of statistically significant association in either cohort.

Our finding of lack of association between family history of bipolar disorder (in first-degree relatives) appears surprising. However, there are several findings from the literature that may help to explain this. The accuracy of proband reports on mental illness has been studied with varying effects seen. Kendler et al. (36) noted that affected co-twins were more likely to report illness in family members than were unaffected co-twins, though this study does not directly address bipolar disorder. In another study that did directly address mania, probands were very good at recognizing history of mania in family members, though still not perfect even for this most visible aspect of bipolar disorder (37). Given the struggles that clinicians themselves have in detecting hypomania, we can expect that family members will similarly struggle. Perhaps most interesting have been the findings that polygenic score does not associate with family history of major mental disorders. In a Danish study that used

registries, thereby avoiding the difficulties associated with proband reports, the polygenic score did not explain genetic variance attributable to family history (38). There are several possibilities for this finding and all are intriguing. Direct first-degree transmission of mental illnesses may have a component of rare variants not found on chips, or variants private to individual families which would not be identified in a population-wide GWAS. Direct family history may be capturing a different type of risk than GWAS, and conversely, the population-wide genetic risk captured by GWAS may appear as a separate portion of the risk seen in close relatives.

Despite the advantages of using a multivariate test for initial hypothesis testing, the finding of significance at one threshold in the NESDA cohort highlights the exploratory nature of such an analysis and the need for careful post-hoc analysis to understand the underlying relationships. The model is designed to allow for the relationships between polygenic loading and individual features to be independent, as one would expect them to be. Therefore, associations with individual features can be in the opposite direction of that hypothesized and the relationship will still contribute to the model. Post-hoc analysis clarified that the major contributor to the model in the NESDA cohort had a relationship in the opposite direction of that seen in STAR*D. Although there was a finding in NESDA with $p < 0.05$, it cannot be seen as replication.

In the context of failure to replicate findings from the primary analysis in additional cohorts, it is important to consider that the initial findings may indeed themselves be due to type I error. However, we must also consider the limitations of our replication efforts. The sample size in both replication cohorts was smaller than in the primary cohort. Another concern is the differences in assessment and ascertainment that are inherent to studies such as these. The most significant ascertainment difference is that the Mannheim cohort consisted entirely of subjects receiving inpatient psychiatric treatment for their depression, while the STAR*D and NESDA cohort were both recruited from a combination of primary care and tertiary care outpatient centers. Even features that appear to be clearly comparable can still be subject to differences in assessment. Age at onset and number of episodes were defined by the physician assessment in STAR*D. Both of these features are being determined retrospectively at a time when the patient is presenting for treatment. In NESDA, these variables were determined by the standardized CIDI psychiatric interview. In the Mannheim cohort, age of onset was assessed by the OPCRIT checklist and the number of episodes was assessed by the SCID-I. In the Mannheim cohort, only 7% of subjects had early-onset depression, while 42% in STAR*D and 30% in NESDA had onset at 18 or under. These variations in prevalence are likely to be due to different assessment strategies. Furthermore, the model did need to be narrowed to five features to provide as much confidence as possible that these elements were comparable. The bipolar spectrum disorder phenotype is complex, and while we captured as much as possible in the STAR*D dataset, some elements could not be assessed. In narrowing the model further we are able to test a part of this model, but have lost some of the complexity.

We have used polygenic scores as a measure of genetic loading for bipolar disorder, assessing genetic loading as a predictor for a bipolar-like presentation of major depression. Given the theoretical and empirical associations amongst the features studied, a multivariate

model provides the best approach to analysis of association. While the results are inconclusive in light of lack of replication, the methods demonstrated here may be helpful in elucidating factors underlying clinical heterogeneity in the presentation of psychiatric disorders. To our knowledge, similar methods have not been applied in this context before.

Acknowledgments

The authors would like to express their gratitude to the Psychiatric Genomics Consortium, without the investigators and participants in those studies this work would not be possible. The authors gratefully acknowledge the investigators of NESDA, the Bonn/Mannheim cohort, and the STAR*D trial as well as the patient participants in those studies.

AW was supported by NIMH 5R25MH094612. AW and EBR were supported by National Institutes of Health (NIH) T32MH017119. MR was supported by the German Federal Ministry of Education and Research (BMBF) within the context of the National Genome Research Network plus (NGFNplus), and the MoodS-Net (grant 01GS08147); and was also supported by the 7th framework programme of the European Union (ADAMS project, HEALTH-F4-2009-242257) and the European Community's Seventh Framework Programme (FP7/2007–2013) under grant agreement n°279227. RHP was supported by National Institute of Mental Health (NIMH) MH086026.

Funding support for NESDA was provided by the Netherlands Scientific Organization (ZonMW Geestkracht program, 10–000–002), Centre for Medical Systems Biology (NWO Genomics), the Neuroscience Campus Amsterdam (NCA) and the EMGO+ institute, and matching funds from participating institutes in NESDA and NTR. Genotyping was funded by the Genetic Association Information Network (GAIN) of the Foundation for the US National Institutes of Health. Genotype data were obtained from dbGaP (<http://www.ncbi.nlm.nih.gov/dbgap>, accession number phs000020.v1.p1).

References

1. Kraepelin, E. Manic-depressive insanity and paranoia. Lvingstone, Edinburgh: E&S; 1921. Reprint of 1921 ed. ed
2. Ghaemi SN, Ko JY, Goodwin FK. "Cade's disease" and beyond: misdiagnosis, antidepressant use, and a proposed definition for bipolar spectrum disorder. *Can J Psychiatry*. 2002; 47:125–134. [PubMed: 11926074]
3. Perlis RH, Brown E, Baker RW, Nierenberg AA. Clinical features of bipolar depression versus major depressive disorder in large multicenter trials. *Am J Psychiatry*. 2006; 163:225–231. [PubMed: 16449475]
4. Angst J, Azorin JM, Bowden CL, et al. Prevalence and characteristics of undiagnosed bipolar disorders in patients with a major depressive episode: the BRIDGE study. *Arch Gen Psychiatry*. 2011; 68:791–798. [PubMed: 21810644]
5. Olfson M, Das AK, Gameroff MJ, et al. Bipolar depression in a low-income primary care clinic. *Am J Psychiatry*. 2005; 162:2146–2151. [PubMed: 16263856]
6. Perlis RH, Smoller JW, Fava M, et al. The prevalence and clinical correlates of anger attacks during depressive episodes in bipolar disorder. *J Affect Disord*. 2004; 79:291–295. [PubMed: 15023510]
7. Deckersbach T, Perlis RH, Frankle WG, et al. Presence of irritability during depressive episodes in bipolar disorder. *CNS Spectr*. 2004; 9:227–231. [PubMed: 14999163]
8. Mitchell PB, Wilhelm K, Parker G, et al. The clinical features of bipolar depression: a comparison with matched major depressive disorder patients. *J Clin Psychiatry*. 2001; 62:212–216. [PubMed: 11305713]
9. Hantouche EG, Akiskal HS. Toward a definition of a cyclothymic behavioral endophenotype: which traits tap the familial diathesis for bipolar II disorder? *J Affect Disord*. 2006; 96:233–237. [PubMed: 16427137]
10. Kochman FJ, Hantouche EG, Ferrari P, et al. Cyclothymic temperament as a prospective predictor of bipolarity and suicidality in children and adolescents with major depressive disorder. *J Affect Disord*. 2005; 85:181–189. [PubMed: 15780688]

11. Mitchell PB, Frankland A, Hadzi-Pavlovic D, et al. Comparison of depressive episodes in bipolar disorder and in major depressive disorder within bipolar disorder pedigrees. *Br J Psychiatry*. 2011; 199:303–309. [PubMed: 21508436]
12. Perlis RH, Uher R, Ostacher M, et al. Association between bipolar spectrum features and treatment outcomes in outpatients with major depressive disorder. *Arch Gen Psychiatry*. 2011; 68:351–360. [PubMed: 21135313]
13. Sklar P, Ripke S, Scott LJ, et al. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat Genet*. 2011; 43:977–983. [PubMed: 21926972]
14. Ripke S, Sanders AR, Kendler KS, et al. Genome-wide association study identifies five new schizophrenia loci. *Nat Genet*. 2011; 43:969–976. [PubMed: 21926974]
15. Purcell SM, Wray NR, Stone JL, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009; 460:748–752. [PubMed: 19571811]
16. Fava M, Rush AJ, Trivedi MH, et al. Background and rationale for the sequenced treatment alternatives to relieve depression (STAR*D) study. *Psychiatr Clin North Am*. 2003; 26:457–494. [PubMed: 12778843]
17. Penninx BW, Beekman AT, Smit JH, et al. The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods. *Int J Methods Psychiatr Res*. 2008; 17:121–140. [PubMed: 18763692]
18. Rietschel M, Mattheisen M, Frank J, et al. Genome-wide association-, replication-, and neuroimaging study implicates HOMER1 in the etiology of major depression. *Biol Psychiatry*. 2010; 68:578–585. [PubMed: 20673876]
19. Garriock HA, Kraft JB, Shyn SI, et al. A genomewide association study of citalopram response in major depressive disorder. *Biol Psychiatry*. 2010; 67:133–138. [PubMed: 19846067]
20. Boomsma DI, Willemsen G, Sullivan PF, et al. Genome-wide association of major depression: description of samples for the GAIN Major Depressive Disorder Study: NTR and NESDA biobank projects. *Eur J Hum Genet*. 2008; 16:335–342. [PubMed: 18197199]
21. Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet*. 2009; 84:210–223. [PubMed: 19200528]
22. Price AL, Patterson NJ, Plenge RM, et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006; 38:904–909. [PubMed: 16862161]
23. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007; 81:559–575. [PubMed: 17701901]
24. Zisook S, Lesser I, Stewart JW, et al. Effect of age at onset on the course of major depressive disorder. *Am J Psychiatry*. 2007; 164:1539–1546. [PubMed: 17898345]
25. Zimmerman M, Mattia JI. A self-report scale to help make psychiatric diagnoses: the Psychiatric Diagnostic Screening Questionnaire. *Arch Gen Psychiatry*. 2001; 58:787–794. [PubMed: 11483146]
26. Rush AJ, Gullion CM, Basco MR, Jarrett RB, Trivedi MH. The Inventory of Depressive Symptomatology (IDS): psychometric properties. *Psychol Med*. 1996; 26:477–486. [PubMed: 8733206]
27. Rush AJ, Trivedi MH, Ibrahim HM, et al. The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biol Psychiatry*. 2003; 54:573–583. [PubMed: 12946886]
28. McGuffin P, Farmer A, Harvey I. A polydiagnostic application of operational criteria in studies of psychotic illness. Development and reliability of the OPCRIT system. *Arch Gen Psychiatry*. 1991; 48:764–770. [PubMed: 1883262]
29. Novick JS, Stewart JW, Wisniewski SR, et al. Clinical and demographic features of atypical depression in outpatients with major depressive disorder: preliminary findings from STAR*D. *J Clin Psychiatry*. 2005; 66:1002–1011. [PubMed: 16086615]

30. Hirschfeld RM, Williams JB, Spitzer RL, et al. Development and validation of a screening instrument for bipolar spectrum disorder: the Mood Disorder Questionnaire. *Am J Psychiatry*. 2000; 157:1873–1875. [PubMed: 11058490]
31. Hirschfeld RM, Holzer C, Calabrese JR, et al. Validity of the mood disorder questionnaire: a general population study. *Am J Psychiatry*. 2003; 160:178–180. [PubMed: 12505821]
32. SAS. SAS v. 9.3. Cary: SAS Publishing; 2011.
33. SAS/STAT 9.3 User's Guide: The GLIMMIX Procedure. Cary: SAS Publishing; 2011.
34. StataCorp. Stata Statistical Software: Release 10. College Station: StataCorp, LP; 2007.
35. Demirkan A, Amin N, Isaacs A, et al. Genetic architecture of circulating lipid levels. *Eur J Hum Genet*. 2011; 19:813–819. [PubMed: 21448234]
36. Kendler KS, Silberg JL, Neale MC, et al. The family history method: whose psychiatric history is measured? *Am J Psychiatry*. 1991; 148:1501–1504. [PubMed: 1928463]
37. Hardt J, Franke P. Validity, reliability and objectivity of the family history method in psychiatry: a meta analysis. *Eur Psychiatry*. 2007; 22:49–58. [PubMed: 17188848]
38. Agerbo E, Mortensen PB, Wiuf C, et al. Modelling the contribution of family history and variation in single nucleotide polymorphisms to risk of schizophrenia: a Danish national birth cohort-based study. *Schizophr Res*. 2012; 134:246–252. [PubMed: 22108675]

Fig. 1.

Variance explained (R^2) by bipolar polygenic risk score at all values of p-value threshold (P_T) in the Sequential Treatment Alternatives to Relieve Depression study (STAR*D) cohort for features which contribute to the multivariate model in that cohort. * $p < 0.05$ in logistic regression, controlling for age, sex, primary versus tertiary care site, and population stratification via principal component analysis.

Tetrachoric correlations between variables predicted to be associated with bipolarity in major depressive disorder

Table 1

	Severity	Early onset	Suicide attempt	Recurrent	Manic symptom	Psychotic symptom	Atypical depression	Family history bipolar
Severity	1							
Early onset	0.18	1						
Suicide attempt	0.31	0.44	1					
Recurrent	0.10	0.66	0.26	1				
Manic symptom	0.04	0.17	0.07	0.09	1			
Psychotic symptom	0.33	0.08	0.10	0.07	0.29	1		
Atypical depression	0.09	0.10	0.01	0.06	0.06	0.24	1	
Family history bipolar	0.05	0.11	-0.04	0.09	0.18	0.04	0.02	1

r* = 0.2 with at least one other variable (in **bold**) was required for inclusion in the multivariate model.

Table 2Post-hoc univariate analyses of the initial STAR*D model^a

Feature	Odds ratio	p-value
Severity	0.98	0.8
Early onset	1.20	0.003 ^b
Recurrent	1.10	0.1
History of suicide attempt	1.21	0.03 ^b
Manic symptom	1.16	0.02 ^b
Psychotic symptom	1.05	0.5
Atypical depression	0.96	0.6

STAR*D = Sequential Treatment Alternatives to Relieve Depression study. Logistic regression adjusting for age, sex, primary versus tertiary care site, and population stratification via principal component analysis.

^aThe polygenic risk score for bipolar disorder derived from the Psychiatric Genome-Wide Association Study Consortium bipolar disorder analysis was used to predict a multivariate model containing these seven features.

^bp < 0.05.

Table 3

Association of bipolar polygenic risk score with the multivariate model containing severity, history of suicide attempt, early onset, recurrent, and atypical depression

P _T	STAR*D (n = 1,274)		NESDA (n = 992)		Mannheim (n = 583)	
	F-statistic	p-value	F-statistic	p-value	F-statistic	p-value
0.001	2.11	0.06	0.5	0.8	0.38	0.9
0.01	2.21	0.05	2.69	0.02	0.78	0.6
0.05	2.52	0.03	2.55	0.03	1.14	0.3
0.1	2.64	0.02	2.02	0.07	0.66	0.7
0.2	2.57	0.03	1.94	0.08	0.78	0.6
0.3	2.61	0.02	1.97	0.08	0.52	0.8
0.4	2.45	0.04	1.76	0.1	0.45	0.8
0.5	2.28	0.04	1.62	0.2	0.45	0.8

STAR*D = Sequential Treatment Alternatives to Relieve Depression study; NESDA = The Netherlands Study of Depression and Anxiety; P_T = p-value thresholds. Generalized linear mixed models with F-statistic for each of eight p-value thresholds in the Psychiatric Genome-Wide Association Study Consortium bipolar analysis results.

Table 4

Association of bipolar polygenic risk score with subclinical mania in each of the three cohorts as defined in the cohorts using the PDSQ, MDQ, and OPCRIT, respectively.

P _T	STAR*D (n = 1,274)		NESDA (n = 992)		Mannheim (n = 583)	
	OR	p-value	OR	p-value	OR	p-value
0.001	1.04	0.5	1.04	0.79	0.97	0.9
0.01	1.18	0.006	1.06	0.65	0.93	0.7
0.05	1.14	0.03	1.3	0.048	1.08	0.7
0.1	1.16	0.02	1.25	0.09	1.06	0.7
0.2	1.14	0.03	1.14	0.32	1.21	0.3
0.3	1.14	0.03	1.08	0.54	1.11	0.5
0.4	1.16	0.02	1.1	0.46	1.06	0.7
0.5	1.15	0.02	1.1	0.49	1.05	0.8

PDSQ = Psychiatric Diagnostic Screening Questionnaire; MDQ = Mood Disorder Questionnaire; OPCRIT = Operational Criteria Checklist; STAR*D = Sequential Treatment Alternatives to Relieve Depression study; NESDA = The Netherlands Study of Depression and Anxiety; P_T = p-value threshold; OR = odds ratio. Logistic regression analysis adjusting for age, sex, and population stratification via principal component analysis.