Evaluation of potential novel variations and their interactions related to bipolar disorders: analysis of genome-wide association study data

Cengizhan Acikel¹
Yesim Aydin Son²
Cemil Celik³
Husamettin Gul⁴
¹Department of Biostatistics, Gulhane Military Medical Academy, ²Department of Health Informatics, Graduate School of Informatics, Middle East Technical University, ³Department of Medical Psychiatry, ⁴Department of Medical Informatics, Gulhane Military Medical Academy, Ankara, Turkey

Background: Multifactor dimensionality reduction (MDR) is a nonparametric approach that can be used to detect relevant interactions between single-nucleotide polymorphisms (SNPs). The aim of this study was to build the best genomic model based on SNP associations and to identify candidate polymorphisms that are the underlying molecular basis of the bipolar disorders.

Methods: This study was performed on Whole-Genome Association Study of Bipolar Disorder (dbGaP [database of Genotypes and Phenotypes] study accession number: phs000017.v3.p1) data. After preprocessing of the genotyping data, three classification-based data mining methods (ie, random forest, naïve Bayes, and k-nearest neighbor) were performed. Additionally, as a nonparametric, model-free approach, the MDR method was used to evaluate the SNP profiles. The validity of these methods was evaluated using true classification rate, recall (sensitivity), precision (positive predictive value), and F-measure.

Results: Random forests, naïve Bayes, and k-nearest neighbors identified 16, 13, and ten candidate SNPs, respectively. Surprisingly, the top six SNPs were reported by all three methods. Random forests and k-nearest neighbors were more successful than naïve Bayes, with recall values >0.95. On the other hand, MDR generated a model with comparable predictive performance based on five SNPs. Although different SNP profiles were identified in MDR compared to the classification-based models, all models mapped SNPs to the DOCK10 gene.

Conclusion: Three classification-based data mining approaches, random forests, naïve Bayes, and k-nearest neighbors, have prioritized similar SNP profiles as predictors of bipolar disorders, in contrast to MDR, which has found different SNPs through analysis of two-way and three-way interactions. The reduced number of associated SNPs discovered by MDR, without loss in the classification performance, would facilitate validation studies and decision support models, and would reduce the cost to develop predictive and diagnostic tests. Nevertheless, we need to emphasize that translation of genomic models to the clinical setting requires models with higher classification performance.

Keywords: Bipolar disorders, GWAS, MDR, Data Mining, SNP, Decision Support

Background
In the first decade of genome-wide association studies (GWASs), researchers predominantly focused on identifying individual genetic variations associated with different phenotypes.¹ This allowed the application of the candidate gene approach to investigate the molecular basis of a disease. However, this strategy has not been very effective for complex genetic diseases, for which the underlying etiology tends to be polymorphisms scattered throughout the genome.² In addition, univariate analysis, while useful in identifying isolated single-nucleotide polymorphisms (SNPs), cannot...
accurately model interactions between genetic and environmental factors and has limited power in associating such interactions with complex genetic diseases. Therefore, multivariate methods are recommended.

Bipolar disorder, also known as manic-depressive illness, is a psychiatric disorder that causes unusual shifts in mood, energy, and activity levels. This disorder is one of the most common psychiatric disorders worldwide, and results in increased disability-adjusted life years in last decades. Bipolar I disorder is characterized by the occurrence of one or more manic or mixed episodes. Conversely, bipolar II disorder manifests as one or more major depressive episodes accompanied by at least one hypomanic episode.

The heritability of bipolar disorder is estimated to be as high as 0.70 in the literature. Bipolar I concordance rates are ~40% in monozygotic twins and between 0% and 10% in dizygotic twins. The combined concordance rates of unipolar depression and bipolar disorder are 67% and 19%, respectively. Previous genetic studies have described multiple candidate genes in as many as 70 chromosomal regions, according to the OMIM database (Online Mendelian Inheritance in Man, http://www.omim.org/) as of April 2015. However, none of the candidates were identified as causative agents. The most frequently detected associations are found on chromosomes 6q, 8q, and 21q. Recent GWASs also identified additional loci without directly implicating them.

Till date, various data mining techniques have been applied, with varying success, to analyze high-dimensional data generated by GWASs of complex genetic disorders. Models generated by data mining can estimate the predictive power of genotype profiles and classify individuals accordingly. These techniques can be applied at various stages throughout a GWAS to, for example, clean and preprocess data, reduce dimension, execute similarity searches, detect associations, implement pattern-based cluster analysis, analyze pathways, and visualize results and data. Nevertheless, limitations remain. For example, building a precise model would require more computing power. In addition, different methods may identify different SNP profiles for the same disease. Finally, the same data mining approach may generate different models for the same disease depending on the assumptions and the technical parameters. The aim of this study was to build an effective genomic model based on SNP associations and to identify candidate polymorphisms that are the underlying molecular basis of the bipolar disorders.

Methods

Data

Genotyping data were obtained from the Whole-Genome Association Study of Bipolar Disorder (dbGaP [database of Genotypes and Phenotypes] study accession number: phs000017.v3.p1) through controlled access. Whole-genome genotyping data were obtained by Affymetrix AFFY 6.0 platform and the number of oligos was declared as 934,940. The National Institute of Mental Health (NIMH) was founded in 1989 and the Genetics Initiative is a branch of NIMH, which collects genetic data for the analysis of Alzheimer’s disease, schizophrenia, and bipolar disorders. The Bipolar Disorder Genetics Initiative aims to collect nationwide demographic, clinical, and diagnostic data about bipolar disorders. NIMH resources are shared worldwide with investigators who work on etiology of bipolar disorders.

The data sets include genotypes and phenotypes of patients with bipolar disorder only, who comprise a subset of all bipolar cases, excluding patients categorized as having bipolar and related disorders. Patients which gave consent for general research use were used as controls, which included 1,767 control cases (1,081 of European ancestry and 686 of African-American ancestry) for the bipolar study, who were a subset of controls for the schizophrenia study. Control subjects have been and continue to be recruited by Financial Knowledge Network (San Jose, CA, USA), a survey and marketing research company. The controls assessed with Composite Instrument for Diagnostic Interviewing and patients with history of mental disorders were excluded.

Analysis

Most widely used predictive classification algorithms in medical data mining are decision trees, artificial neural networks, rule-based classification methods, support vector machines, random forests, k-nearest neighbors (kNNs), and naive Bayesian methods. These algorithms can also be applied to identify disease-associated SNPs. On the other hand, nonparametric approaches such as combinatorial partitioning, neural networks, and multifactor dimensionality reduction (MDR) detect relevant interactions between SNPs or genes by either reducing the dimensionality of the data or identifying informative hidden patterns. An advantage of these methods is that they do not make assumptions about the relationships between SNPs and phenotypes, and are therefore purely data driven. For example, MDR can use SNP–SNP interactions to predict disease susceptibility. To reduce the dimensionality of
Identification of candidate polymorphisms in bipolar disorders genome-wide association data, MDR attempts to identify combinations of SNPs instead of individual SNPs, and thereby avoids overparameterization, from which traditional regression-based methods suffer.\textsuperscript{25}

The analytical approach is performed as depicted in Figure 1. Genotyping and phenotyping data were first integrated, cleaned, and transformed. Data were then preprocessed in two key steps, using Affymetrix Gene Console\textsuperscript{TM} (Affymetrix Inc., Santa Clara, CA, USA) and PLINK (open-source whole-genome association analysis toolset version 1.8) to filter signals, perform background correction, normalization, and assess data quality. SNPs were filtered based on minor allele frequency (MAF) of <5\% and failure of the Hardy–Weinberg equilibrium test, as indicated by \textit{P}-value of <0.001. Univariate chi-square tests were also performed by PLINK. Subsequently, genome-based probabilistic models of bipolar disorder were generated using random forests, naïve Bayesian, kNNs, and MDR.

Random forests
New methods have been emerging to address the limitations of classical statistics in dealing with highly dimensional data. One of the recent methods is random forests, an ensemble learning method broadly applicable in data mining and machine learning. The technique is nonparametric, tree-based, and combines the concept of nearest neighbors with bagging.\textsuperscript{26} In this approach, one-step-at-a-time node splitting enables trees (and hence forests) to impose regularization and thereby effectively analyze data sets with “large \textit{P} (probability) values and small number (\textit{n}) of case \textit{s}”. In addition, grouping trees based on these properties allows the method to deal with correlated and interacting variables.\textsuperscript{27} We used the randomForest package for \textit{R} from the CRAN repository. To validate results, we used 1,000 trees and at least 100 SNPs at every split in the random forests. The final model is a random forest of numerous decision trees.\textsuperscript{27–31} The most important advantage is that reduction in dimensionality is not required. The algorithm allows the classification and prediction of the effect of SNPs and is significantly more successful than simple decision trees in analyzing gene expression.

Naïve Bayes
Naïve Bayes is another excellent method to analyze genomic data. It is one of the earliest machine-learning methods and is being used for over 50 years in biomedical informatics. The technique is computationally efficient and performs better than expected in classification tasks.\textsuperscript{32} However, miscalibration can be an issue when numerous features are used, and

\textbf{Figure 1. Data analysis flowchart.}

\textbf{Abbreviations:} BDO, bipolar disorder only; MAF, minor allele frequencies; HWE, Hardy–Weinberg equilibrium; kNN, k-nearest neighbor; MDR, multifactor dimensionality reduction; SNP, single-nucleotide polymorphism; PLINK, open-source whole-genome association analysis toolset version 1.8.
it tends to make predictions with posterior probabilities too close to 0 or 1. Nevertheless, the technique is successfully applied to diagnose diseases, identify news articles of interest, classify web pages by topic, and assign proteins to functional families. In this study, we used the CRAN e1071 package for R to perform naïve Bayes modeling without double controlling Laplace smoothing. Thus, the epsilon range to apply Laplace smoothing, as well as the threshold for replacing cells with probabilities within the epsilon range, was not defined.

K-nearest neighbors

Unlike the other classification approaches, kNN does not build a classifier using the training data. Instead, it searches for k data points closest to the test object and uses the features of these neighbors to classify the new object. In instances where multiple classifications are possible, vote counting is applied. We used the kNN package for R from the CRAN repository, with number of neighbors set to k=5. The minimum number of votes required for a definite decision was not defined, and the “frequency of the majority class” was returned as the attribute probability.

Multifactor dimensionality reduction

One of the biggest challenges in GWASs is to evaluate SNP–SNP or gene–gene interactions. While logistic regression can be used to analyze such interactions, overfitting appears to be a significant issue. MDR, a nonparametric, model-free method, is an alternative approach. MDR basically reduces multilocus genotypes into high- or low-risk groups, based on the number of affected and control cases present in a group. Groups in which the ratio between affected and control cases exceeds a predefined threshold are considered as high-risk groups. The method is more powerful than logistic regression in testing high-order interactions and has many possible variations, including generalized MDR. To perform MDR, we used the MDR package for R from the CRAN repository, and analyzed data with parameters \( K=2, \ cv=5, \) ratio = NULL, equal = “HR”, and genotype = c(0, 1, 2). The ratio of MDR was assumed to be 1:100.

Model building

All four models were built and evaluated as described in the following steps: first, subsets of features were generated automatically using the Bioconductor package in R and evaluated using distance measures. Models were then run until the first 50, 100, or 150 SNPs were selected. To validate the results, 1,000 trees and at least 100 SNPs were used at every split in random forests. In naïve Bayes and kNNs, 80% of the randomly selected cases were used for the training set, while the remaining 20% were used as the test set. In MDR, models were evaluated by fivefold cross-validation. Implicated procedures are summarized in a flowchart and presented in Figure 1.

Results

Preprocessing

A total of 2,371 individual genotypes were analyzed, of which 604 were bipolar patients and 1,767 were controls. The distribution of cases into consent groups was done as reported in Table 1.

Random forests, naïve Bayes, and kNNs identified 16, 13, and ten candidate SNPs, respectively. The top six SNPs identified were common in all the three methods (Table 2).

Comparison of different models

Random forests and kNNs were more successful than naïve Bayes, with recall values above 0.95 (Table 3). On the other hand, MDR generated a model with comparable predictive performance based on only five SNPs identified by the analysis of two-way and three-way interactions.

### Table 1 Main descriptive statistics

<table>
<thead>
<tr>
<th>Group</th>
<th>Race</th>
<th>Frequency</th>
<th>Valid percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDO</td>
<td>Total</td>
<td>604</td>
<td>34.2</td>
</tr>
<tr>
<td>GRU</td>
<td>Total</td>
<td>1,767</td>
<td>65.6</td>
</tr>
<tr>
<td>BDO</td>
<td>EA</td>
<td>339</td>
<td>56.1</td>
</tr>
<tr>
<td></td>
<td>AA</td>
<td>265</td>
<td>43.9</td>
</tr>
<tr>
<td>GRU</td>
<td>EA</td>
<td>1,081</td>
<td>61.2</td>
</tr>
<tr>
<td></td>
<td>AA</td>
<td>686</td>
<td>38.8</td>
</tr>
</tbody>
</table>

**Abbreviations:** GRU, Control genotypes with General research use consent; BDO, Bipolar disorders only; EA, European ancestry; AA, African American ancestry.


Analysis of SNP–SNP interactions

MDR was used to investigate two-way or three-way SNP–SNP interactions, although three-way interactions were favored due to the large number of SNPs. The most significant two-way interactions were between rs17736182 and rs2055710, which map to two genes KLHL1 and DOCK10, respectively. Patients with specific allelic profiles for these SNPs showed the highest risk (67.54%) of having bipolar disorder. Analysis of three-way interactions identified a risk allele for rs2483023, an SNP on the LEMD1 gene, along with two other novel SNPs. In the two-way interaction analysis, the patients carrying allele A for rs17736182 were found to have an equal risk as the patients carrying allele A for rs2055710, which was 67.54%. In the three-way interaction assessment, patients carrying allele C for rs9372649 showed equal risk with the patients carrying allele A for rs12145634 and allele C for rs2483023, with the prediction performance of 77.2%.

Conclusion

Variant calling and annotation is one way to search for SNPs associated with the disease, but this method is mainly biased toward annotated coding regions of the genome. However, hypothesis-free methods, such as presented here, do not rely on prior knowledge or genomic annotation, and therefore present a potential approach to overcome this bias. This study stands out from previous studies as three different classification methods and MDR (as an SNP–SNP interaction model) were compared for the first time in the literature on the same GWAS data, the dbGAP’s bipolar disorders data. As a result, the advantages and disadvantages of each data mining method were assessed. Also, another novelty of the study is that all cases with bipolar-related disorders were filtered, and the data of cases with bipolar disorders only were analyzed.

The random forests were found to have the best overall classification rate, as well as a more straightforward assessment of the classification errors. Indeed, extensive research

Table 2 Single-nucleotide polymorphisms identified in the genome-based model for RF, kNN, and NB methods

<table>
<thead>
<tr>
<th>RS ID</th>
<th>RF</th>
<th>kNN</th>
<th>NB</th>
<th>Multidimensional reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs6785</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2194124</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs4792189</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs7569781</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs9375098</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs10415145</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs10857580</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs11015814</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs11015877</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs732183</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs11023096</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1328392</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2791142</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1861226</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs4654814</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs219506</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2055710</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2483023</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs9372649</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs12145634</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs17736182</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: RF, random forest; NB, naive Bayes; kNN, k-nearest neighbor.

Table 3 Comparison of the performance of the classification-based models with MDR

<table>
<thead>
<tr>
<th>Feature</th>
<th>Method</th>
<th>RF</th>
<th>NB</th>
<th>kNN</th>
<th>MDR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Two-way</td>
</tr>
<tr>
<td>Validity</td>
<td>Classification accuracy</td>
<td>0.734</td>
<td>0.702</td>
<td>0.733</td>
<td>0.647</td>
</tr>
<tr>
<td></td>
<td>F-measure</td>
<td>0.853</td>
<td>0.785</td>
<td>0.841</td>
<td>0.764</td>
</tr>
<tr>
<td></td>
<td>Precision</td>
<td>0.743</td>
<td>0.845</td>
<td>0.754</td>
<td>0.675</td>
</tr>
<tr>
<td></td>
<td>Recall</td>
<td>0.998</td>
<td>0.734</td>
<td>0.954</td>
<td>0.664</td>
</tr>
<tr>
<td>Overfit</td>
<td>Very resistant since boot strap selection is performed</td>
<td>Relatively risky</td>
<td>Bootstrapping performed to avoid overfit</td>
<td>Risky k-fold cross-validation used to overcome overfit problem</td>
<td></td>
</tr>
<tr>
<td>Advantages</td>
<td>Nonparametric interpretable</td>
<td>Resistant to noise</td>
<td>Good for eliminating missing values</td>
<td>Simple, flexible Nonparametric test</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Resistant to noise</td>
<td>Sensitive to inconsistent data</td>
<td>Accuracy degraded by correlated variables</td>
<td>Arbitrary decision boundaries</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nondeterministic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: RF, random forest; NB, naive Bayes; kNN, k-nearest neighbor; MDR, multifactor dimensionality reduction.
using actual or simulated data has demonstrated decision
trees to be very flexible and easy to debug. However,
simple decision trees tend to overfit the data more than
other techniques. Therefore, researchers generally prune
trees and tune procedures to do so. Random forest method
was originally developed to overcome this issue and, this
study has generated the most accurate classification without
overfitting. The performance of kNNs was comparable.
However, naïve Bayes had the lowest classification rate with
regard to tabular data simulation, although application of
the feature selection option may improve its performance.\(^4\)
Previous studies, which compare random forests with other
classification methods, also agree with our observation. For
instance, Lunetta et al\(^2, 3\) conducted a simulation experiment
to evaluate the ability of random forests to detect interacting
SNPs and found that it outperformed Fisher’s exact test, even
though both methods were comparable in the absence of such
interactions. One major concern about random forest is its
high computational cost compared to the others discussed
here. In addition, random forest has been reported to be very
sensitive to noise or unbalanced data sets, whereas k-nearest
neighbor algorithms are more efficient and stable.\(^4\)
Overall classification rates were between 0.70 and 0.75. Due to
bootstrapping and the nature of the methods used, no model
achieved classification rates better than 0.9.

In addition, MDR method revealed different SNPs
through analyses of two-way and three-way interactions. This
may be due to the increased frequency of SNP interactions in
polygenic diseases. Nevertheless, the classification success
of MDR, based on only three SNPs, was comparable to other
models. Notably, physical and functional annotation of the
SNPs showed one SNP mapping to the \(DOCK10\) gene, which
is also identified with other three methods investigated here.
The remaining two SNPs were mapped to the genes related to
those found by other models, although these genes were
not common to all. So, the high classification performance
and relevant biological annotation of the SNPs discovered
support that MDR would be an effective alternative method
to evaluate SNP–SNP interactions. Also, the reduced number
of SNPs, without loss in classification performance, would
facilitate validation studies and decision support models, and
would reduce the cost to develop predictive and diagnostic
tests. Nevertheless, we acknowledge that translation of
genomic models to the clinical setting will require models
with higher classification performance.

Previous analyses of bipolar disorder genotypes revealed
a high level of complexity, and a consensus profile of
associated SNPs or genes could not be identified in those
studies. Here, six candidate SNPs associated with bipolar
disorders were common in all three classification-based
methods. These SNPs identified two candidate genes that may
potentially be the causative agents. Indeed, pathway analysis
in GeneMANIA (http://genemania.org) indicates that these
genes are closely associated with psychiatric disorders.\(^4\)
Even though the classification performance is not sufficient
for translation of the findings into a clinical diagnostic test,
we suggest that the consensus SNP profile obtained from the
three classification-based methods has high potential to be
the causative variants, and further experimental validation
would be productive.

**Acknowledgments**

Funding support for the Whole-Genome Association
Study of Bipolar Disorder was provided by NIMH, and the
genotyping of samples was provided through the Genetic
Association Information Network. The data sets used for
the analyses described in this manuscript were obtained
from the database of Genotypes and Phenotypes (dbGaP)
through dbGaP accession number phs000017.v3.p1. Samples and associated
phenotype data for the Collaborative Genomic Study of Bipolar Disorder were provided by the NIMH Genetics
Initiative for Bipolar Disorder. Data and biomaterials were
collected in four projects that participated in NIMH Bipolar
Disorder Genetics Initiative. From 1991–1998, the principal
investigators and coinvestigators were: Indiana University,
Indianapolis, IN, USA, U01 MH46282, John Nurnberger,
MD, PhD, Marvin Miller, MD, and Elizabeth Bowman, MD;
Washington University, St Louis, MO, USA, U01 MH46280,
Theodore Reich, MD, Allison Goate, PhD, and John Rice,
PhD; Johns Hopkins University, Baltimore, MD, USA, U01
MH46274, J Raymond DePaulo, Jr, MD, Sylvia Simpson,
MD, MPH, and Colin Stine, PhD; NIMH Intramural Research
Program, Clinical Neurogenetics Branch, Bethesda, MD,
USA, Elliot Gershon, MD, Diane Kazuba, BA, and Elizabeth
Maxwell MSW. Data and biomaterials were collected as
part of ten projects that participated in the NIMH Bipolar
Disorder Genetics Initiative. From 1999–2003, the principal
investigators and coinvestigators were: Indiana University,
Indianapolis, IN, R01 MH59545, John Nurnberger, MD,
PhD, Marvin J Miller, MD, Elizabeth S Bowman, MD, N
Leela Rau, MD, P Ryan Moe, MD, Nalini Samavedy,
MD, Rif El-Mallakh, MD (at University of Louisville),
Husseini Manji, MD (at Wayne State University), Debra A

---

**References**

1. Husseini Manji, MD (at Wayne State University), Debra A

---

**Dovepress**

Neuropsychiatric Disease and Treatment 2016:12

submit your manuscript | [www.dovepress.com](http://www.dovepress.com)

**Dovepress**

Neuropsychiatric Disease and Treatment 2016:12
Glitz, MD (at Wayne State University), Eric T Meyer, MS, Carrie Smiley RN, Tatiana Foroud, PhD, Leah Flury, MS, Danielle M Dick, PhD, Howard Edenberg, PhD; Washington University, St Louis, MO, USA, R01 MH059534, John Rice, PhD, Theodore Reich, MD, Allison Goate, PhD, Laura Bierut, MD; Johns Hopkins University, Baltimore, MD, USA, R01 MH59533, Melvin McInnis MD, J Raymond DePaulo, Jr, MD, Dean F MacKinnon, MD, Francis M Mondimore, MD, James B Potash, MD, Peter P Zandi, PhD, Dimitrios Avramopoulos, and Jennifer Payne; University of Pennsylvania, PA, USA, R01 MH59553, Wade Berrettini, MD, PhD; University of California, Irvine, CA, USA, R01 MH60068, William Byerley, MD, and Mark Vawter, MD; University of Iowa, IA, USA, R01 MH059548, William Coryell, MD and Raymond Crowe, MD; University of Chicago, IL, USA, R01 MH59353, Elliot Gershon, MD, Judith Badner, PhD, Francis McMahon, MD, Chunyu Liu, PhD, Alan Sanders, MD, Maria Caserta, Steven Dinwiddie MD, Tu Nguyen, Donna Harakal; University of California, San Diego, CA, USA, R01 MH59567, John Kelsoe, MD, Rebecca McKinney, BA; Rush University, IL, USA, R01 MH059556, William Scheftner, MD, Howard M Kravitz, DO, MPH, Diana Marta, BS, Annette Vaughn-Brown, MSN, RN, and Laurie Bederow, MA; NIMH Intramural Research Program, Bethesda, MD, USA, 1Z01MH002810-01, Francis J McMahon, MD, Layla Kassem, PsyD, Sevilla Detera-Wadleigh, PhD, Lisa Austin, PhD, Dennis L Murphy, MD.

This study performed as a part of PhD thesis in Department of Health Informatics, Graduate School of Informatics, Middle East Technical University, Ankara, Turkey.

Disclosure

The authors report no conflicts of interest in this work.

References

3. Eichler EE, Flint J, Gibson G, et al. Missing heritability and strategies of Health Informatics, Graduate School of Informatics, Wadleigh, PhD, Lisa Austin, PhD, Dennis L Murphy, MD.


