BDgene: A Genetic Database for Bipolar Disorder and Its Overlap With Schizophrenia and Major Depressive Disorder

Su-Hua Chang, Lei Gao, Zhao Li, Wei-Na Zhang, Yang Du, and Jing Wang

Background: Bipolar disorder (BD) is a common psychiatric disorder with complex genetic architecture. It shares overlapping genetic influences with schizophrenia (SZ) and major depressive disorder (MDD). Large numbers of genetic studies of BD and cross-disorder studies between BD and SZ/MDD have accumulated numerous genetic data. There is a growing need to integrate the data to provide a comprehensive data set to facilitate the genetic study of BD and its highly relevant diseases.

Methods: BDgene database was developed to integrate BD-related genetic factors and shared ones with SZ/MDD from profound literature reading. On the basis of data from the literature, in-depth analyses were performed for further understanding of the data, including gene prioritization, pathway-based analysis, intersection analysis of multidisease candidate genes, and pathway enrichment analysis.

Results: BDgene includes multiple types of literature-reported genetic factors of BD with both positive and negative results, including 797 genes, 3119 single nucleotide polymorphisms, and 789 regions. Shared genetic factors such as single nucleotide polymorphisms, genes, and regions from published cross-disorder studies among BD and SZ/MDD were also presented. In-depth data analyses identified 43 BD core genes; 70 BD candidate pathways; and 127, 79, and 107 new potential cross-disorder genes for BD-SZ, BD-MDD, and BD-SZ-MDD, respectively.

Conclusions: As a central genetic database for BD and the first cross-disorder database for BD and SZ/MDD, BDgene provides not only a comprehensive review of current genetic research but also high-confidence candidate genes and pathways for understanding of BD mechanism and shared etiology among its relevant diseases. BDgene is freely available at http://bdgene.psych.ac.cn.

Key Words: BDgene, bipolar disorder, data analysis, genetic database, genetic overlap, major depressive disorder, schizophrenia

Bipolar disorder (BD) is a common and severe psychiatric disorder characterized by the cycles between bouts of mania and depression (1). The lifetime prevalence of BD is estimated to be between 1% and 2% (2,3). With the extensive impairment and high risk of suicide (1,4), BD causes a significant impact on patients' quality of life, as well as a considerable economic burden on families and society (5). The etiologic mechanisms for BD are not well understood, but empirical data consistently suggest the polygenic character of BD with estimated heritability ranging from 80% to 85% (6). Meanwhile, several BD clinical features, including psychosis and suicidality, can also be observed in schizophrenia (SZ) and major depressive disorder (MDD) (7). Increasing evidence has indicated that familial coaggregation or comorbidity between these disorders is mainly attributable to overlapping genetic influences (7–10). These findings have raised questions on how these disorders are etiologically connected. To unveil the disease mechanism of BD and its mutual pathogenesis with SZ/MDD, it is of vital importance to study the genetic basis of BD and its overlap with SZ and MDD.

During the past decade, large numbers of association and linkage studies and meta-analyses aiming to explore genetic susceptibility of BD have been conducted, and numerous susceptibility variants, genes, and chromosomal regions have been reported to be associated with BD (11,12). However, these results are scattered in numerous publications and are often inconsistent. For example, several studies supported the association of gene *RGS4* (regulator of G-protein signaling 4) with BD using case–control or family-based association study design (13–15). However, some studies with similar design did not detect the association between tagging single nucleotide polymorphisms (SNPs) in *RGS4* and BD (16,17). These scattered and inconsistent results have made it difficult for researchers to acquire a global understanding of all positive and negative findings. Therefore, there is a growing need to integrate genetic data of BD from various genetic studies to present a systematic review of current genetic research on BD. In the meantime, more and more cross-disorder studies have tried to investigate the genetic overlapping among BD, SZ, and MDD (7,18), which also calls for a systematic integration of corresponding results to lay a solid foundation for discovery of shared mechanism of BD and its highly relevant diseases. In addition, based on integration of data from literature, in-depth analysis of published data will help to provide reliable guide for experimental verification, establish a connection among different types of data, and stimulate novel academic perspectives. For example, the large numbers of genetic data have brought challenge for selecting high-confidence candidate genes for experimental verification. Gene prioritization analysis is proposed to tackle the challenge by ranking genes according to disease relevance (19). In addition, as a complex disease, BD may

From the Key Laboratory of Mental Health, Institute of Psychology, Chinese Academy of Sciences (SHC, LG, ZL, WNZ, YD, JW); and University of Chinese Academy of Sciences (LG, ZL, WNZ, YD), Beijing, China.

Authors SHC, LG, ZL, and WNZ contributed equally to this work.

Address correspondence to Jing Wang, Ph.D., Key Laboratory of Mental Health, Institute of Psychology, Chinese Academy of Sciences, 16 Lincui Road, Chaoyang District, Beijing 100101, China; E-mail: wangjing@psych.ac.cn.

Received Dec 17, 2012; revised Mar 27, 2013; accepted Apr 12, 2013.

0006-3223/$36.00
http://dx.doi.org/10.1016/j.biopsych.2013.04.016

result from multiple genes that disrupt one or more pathways (20). The traditional association approach examines individual SNPs/genes and ignores their combined effects (21). Pathway-based analysis (PBA) is an effective method to explore the combined effects of multiple genes by detecting disease-related pathways from genome-wide association study (GWAS) data (22).

We developed BDgene, a genetic database aiming to integrate multitype genetic factors of BD from published genetic studies. The genetic factors not only include variants (e.g., SNP, haplotype, copy number variation [CNV], and other variants), genes, and regions but also gene–gene interactions and pathways. In addition, BDgene emphasizes the overlapping genetic factors between BD and SZ/MDD from cross-disorder studies for shared disease etiology research. In-depth data analyses were performed, and the results were presented in BDgene, including gene prioritization analysis, pathway-based analysis for GWAS data, intersection analysis of multidisease candidate genes, functional annotation and pathway enrichment analysis for BD core genes and BD shared genes with SZ/MDD. BDgene is targeted to help unveil the genetic basis of BD and its shared mechanism with SZ/MDD.

Methods and Materials

Literature Search and Data Extraction

To obtain the literature-origin BD-related genetic factors, a comprehensive search of BD-related genetic publications in PubMed (http://www.ncbi.nih.gov/pubmed) was made by using the following search terms: *("bipolar" [Title/Abstract] OR "manic depressive" [Title/Abstract] OR "manic depression" [Title/Abstract]) AND (polymorphism [Title/Abstract] OR SNP [Title/Abstract] OR haplotype [Title/Abstract] OR interaction [Title/Abstract] OR variant [Title/Abstract] OR variation [Title/Abstract] OR mutation [Title/Abstract] OR CNV [Title/Abstract] OR "copy number variation" [Title/Abstract] OR repeats [Title/Abstract] OR deletion [Title/Abstract] OR duplication [Title/Abstract] OR ((gene [Title/Abstract] OR locus [Title/Abstract] OR chromosome [Title/Abstract] OR genetic [Title/Abstract] OR genome [Title/Abstract] OR genomic [Title/Abstract]) AND (linkage [Title/Abstract] OR associat* [Title/Abstract])))*. It resulted in 4836 English publications as of March 1, 2013. Abstracts of these publications were manually screened on the basis of inclusion criteria (i.e., genetic studies [association, linkage, and genetic interaction studies] using study designs of family-based, case–control, pedigree, twin, or affected sib pairs, and with the target of identifying genetic susceptibility factors of BD in diagnosed patients) were included. The following studies were excluded: 1) publications about pharmacology, sociology, electrophysiology, neurophysiology, and behavioral research, which are not genetic studies or do not focus on genetic susceptibility of BD; 2) animal model research; 3) review articles without data analysis and new statistical results; and 4) genetic studies only focus on internal clinical variables, such as different age of onset, suicidal behavior, and medicine response, but do not study BD as one phenotype to identify the disease susceptibility. After filtering, 789 articles were retained. Furthermore, to analyze disease-related pathways, there were seven studies about PBA for BD GWAS data collected (20,23–28). In all, 796 studies were included in BDgene.

The full text of each eligible publication was read carefully, and detailed information of each genetic factor in the study was extracted manually, including allele change, statistical values (p value, odds ratio, etc.), and author comments. For better interpretation of the results, study design, sample population, sample size, analytical method, as well as result summary from each study were presented. For PBA studies, all pathways reported in PBA articles were collected, and the analyzed data set, specific PBA approach, and detailed parameters were also provided. Meanwhile, to illustrate the association between genetic candidates and disease, all statistical results from the original publications were categorized into "Positive," "Negative," or "Trend" according to the criteria described in our previous studies (29): 1) for candidate–gene association studies, the result with $p < .05$ was defined as "Positive." 2) For GWAS, $p < 1 \times 10^{-8}$ suggested a Positive result, $p > 1 \times 10^{-5}$ suggested a Negative result, and a value falling between these thresholds represented a Trend result. 3) Mutational result was classified as Trend if no statistical significance was presented. 4) For linkage studies, significance levels of LOD (logarithm [base 10] of odds) > 3.3, $1.9 < \text{LOD} < 3.3$ and $\text{LOD} < 1.9$ were used for Positive, Trend, and Negative results respectively as proposed by Lander and Kruglyak (30). If other statistical values were used, the criteria were referred to the statistical method in original papers.

In addition, the genetic studies that investigated the association of genetic factors with more than one disease in one study under identical or similar methodologic conditions were regarded as "cross-disorder studies." We focused on genetic cross-disorder studies on the two (BD-SZ, BD-MDD) or three (BD-SZ-MDD) diseases. Among 796 studies collected in BDgene, 279 articles were cross-disorder studies. For these, detailed statistical results and author comments for each genetic factor in BD and SZ and/or MDD were extracted apart from the original sample information. Statistical result of each genetic factor in each disease was also categorized into Positive, Negative, or Trend by using the same criteria as described above.

Data Analysis

On the basis of the data from the literature, in-depth data analyses were implemented as described in this section (Figure 1).

Gene Prioritization. With the purpose of helping researchers select the most promising genes from large list of candidate genes for further verification study, gene prioritization analysis was implemented to prioritize 797 BD candidate genes hosted in BDgene by adopting five multiple-source-based gene prioritization tools, i.e., Endeavour (31,32), DIR (33), ToppGene (34), ToppNet (34), and TargetMine (35). Detailed descriptions on selection

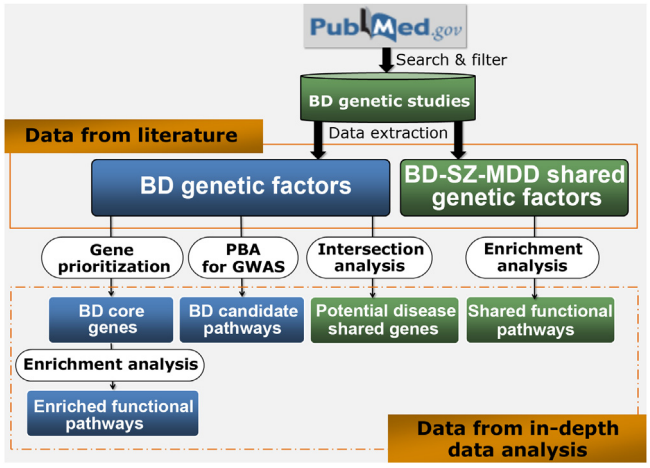

Figure 1. The process of data collection and data analysis in BDgene. BD, bipolar disorder; GWAS, genome-wide association; MDD, major depressive disorder; PBA, pathway-based analysis; SZ, schizophrenia.