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

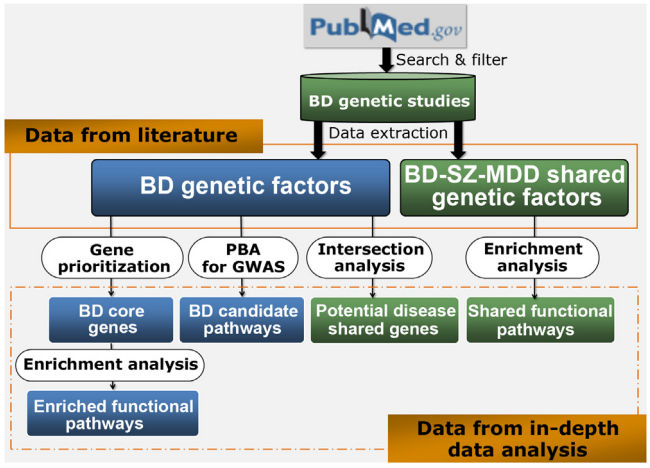

Figure 1. The process of data collection and data analysis in BDgene. BD, bipolar disorder; GWAS, genome-wide association; MDD, major depressive disorder; PBA, pathway-based analysis; SZ, schizophrenia.

criteria of tools for analysis and the analytical procedures were described in our previous publication on attention-deficit/hyperactivity disorder (36). Briefly, most gene prioritization tools are designed based on the principle that the most promising candidates will be those that are similar to known genes that already have solid evidence of involvement in disease susceptibility (19). These known genes were called "training genes." In our analysis, 20 genes were included in the training gene set (Table S1 in Supplement 1), among which 18 were extracted from an extensively cited article about BD liability genes (11) with a threshold of at least three positive results in BDgene, and the other two important genes, *ANK3* and *CACNA1C*, were reported and replicated by several GWASs and meta-analyses (37–39). All candidate genes hosted in BDgene were analyzed by the five tools, and each tool generated a prioritized gene list. To reduce false positives, only those genes predicted by at least three of the five tools were regarded as the most promising candidate genes (called "prioritized genes"). Both training genes and prioritized genes were defined as BD "core genes."

Pathway-Based Analysis for GWAS. In addition to the pathways collected from the seven published BD PBA papers, which analyzed five GWAS data sets in total, to capture more candidate pathways related with BD, we performed pathway-based analysis for 12 available independent BD GWAS data sets by using an improved gene set enrichment analysis (*i*-GSEA) method developed by our group (28). The 12 BD GWAS data sets included two data sets from authors (40,41), two from the Database of Genotypes and Phenotypes (dbGaP) (42), one from the Wellcome Trust Case Control Consortium (43), and seven from the Psychiatric GWAS Consortium of National Institute of Mental Health (https://www.nimhgenetics.org). For raw genotype data from Wellcome Trust Case Control Consortium and National Institute of Mental Health, Cochran-Armitage trend test was used to create a p value list of SNPs by using PLINK (44). Annotated Gene Ontology (GO) terms downloaded from Molecular Signatures Database (MSigDB) v3.0 (45) and pathways from Kyoto Encyclopedia of Genes and Genomes (KEGG) (46) and BioCarta (http://www.biocarta.com) were used as reference gene sets. To map SNPs to genes, "5 kb upstream and downstream of gene" was used (47). Other parameters were referred to the default parameters in *i*-GSEA online tool *i*-GSEA4GWAS (28). Statistically significant pathways/gene sets with false discovery rate $<.05$ were included in BDgene.

Intersection Analysis of Multidisease Candidate Genes. Besides the shared genes from published cross-disorder studies between BD and SZ/MDD, the identical genes reported to be associated with BD and SZ/MDD in separate studies might also provide clues to a disease shared mechanism. SZGene is a collection of SZ-related genes and polymorphisms from published genetic association studies (48), and MK4MDD is a multilevel knowledge base for MDD that provides a list of MDD-related genes from published genetic studies (49). To provide new candidate genes for cross-disorder research in the future, gene lists from SZGene, MK4MDD, and BDgene were compared in which the intersection of genes from BDgene-SZGene and BDgene-MK4MDD were recognized as potential shared genes for BD-SZ and BD-MDD, respectively. Among all overlapping BD-SZ shared genes and BD-MDD shared genes (including from both cross-disorder studies on the two diseases and intersection analysis), the remaining genes after excluding those with evidence from cross-disorder studies on the three diseases (BD-SZ-MDD) were considered potential BD-SZ-MDD shared genes from intersection analysis. Gene descriptions and resulting categories from SZGene and MK4MDD were also provided.

Functional Annotation and Pathway Enrichment Analysis. To facilitate users' understanding of the function of all reported candidate SNPs and genes, functional annotation for them was made. SNP functions were annotated as "synonymous," "non-synonymous," "frame shift," and so on by using the data from Ensembl (47). In addition, we conducted linkage disequilibrium (LD) analysis for the literature reported SNPs in BDgene to capture more functional candidates, especially candidate causal SNPs. The LD data were downloaded from HapMap FTP (file transfer protocol), which were compiled from merged genotype data from phases I+II+III (HapMap rel #27, National Center for Biotechnology Information B36) for markers up to 200 kb apart (50). The population used in the LD analysis was consistent with original studies. SNPs in high LD with the literature-origin SNP ($r^2 > .8$) were selected as LD proxies (51). Functional annotation on genes included gene-related GO (52) terms, gene-related pathways from KEGG (46) and BioCarta (http://www.biocarta.com), and protein–protein interactions from Human Protein Reference Database (53). Furthermore, to interpret the function of BD core genes and literature-reported shared genes with at least one positive result, pathway enrichment analysis was implemented by inputting the gene lists into DAVID 6.7 to highlight the most relevant pathways associated with the gene lists (54,55).

Results

On the basis of a comprehensive literature review and in-depth data analysis, BDgene not only contains a genetic data set for BD and the shared ones with SZ and MDD that were published so far but also provides a series of in-depth analytical results. All data from BDgene were stored and managed in a MySQL relational database, and a user-friendly web interface was developed to help users search and use the data.

BD Genetic Factors

Through depth-mining of 796 papers, BDgene contains 797 candidate genes for BD, of which 447 genes have at least one positive finding according to our categorization of statistical results. There are 45 genes examined by at least five studies in BDgene, in which *SLC6A4*, *BDNF*, and *DRD2* are the top three studied genes reported by 41, 35, and 26 publications, respectively. In addition to genes, BDgene also provides multitype genetic factors including 3119 SNPs (882 of them showed significant associations with BD in at least one study), 1473 haplotypes, 401 CNVs, 674 other variants (e.g., variable number of tandem repeats, microsatellite, duplication, insertion/deletion), 38 pairs of gene–gene interactions, and 789 chromosomal regions (92 positive regions from linkage studies). Moreover, to explore the combined effect of genes, BDgene contains 245 pathways identified by pathway-based analysis for BD GWAS data, among which 188 pathways were obtained from published BD PBA papers, and 70 pathways were acquired from our PBA by implementing *i*-GSEA for 12 BD GWAS data sets. Remarkably, 13 of these pathways were identified by both published PBA papers and our own PBA analysis, including "calcium channel activity" (GO:0005262) and "central nervous system development" (GO:0007417). Full statistics for BD-related multitype genetic factors are shown in Table 1.

BD Shared Genetic Factors With SZ and MDD

BDgene contains 285 shared genes for BD-SZ, 120 for BD-MDD, and 49 for BD-SZ-MDD from published cross-disorder studies, in which 83, 47, and 10 genes have at least one positive

Table 1. Data Content and Statistics of Bipolar Disorder Genetic Factors in BDgene (Through March 1, 2013)

Data Type	Data Statistics[a]
Variant	
SNP	3119 (882)
Haplotype	1473 (655)
CNV	401 (NA)
Other[b]	674 (191)
Gene	797 (447)
Gene–Gene Interaction	38 (26)
Region	789 (92)
Pathway[c]	245
From PBA Papers	188
From *i*-GSEA4GWAS Analysis[d]	70

CNV, copy number variation; *i*-GSEA4GWAS, improved gene set enrichment analysis method online tool; NA, not applicable; PBA, pathway-based analysis; SNP, single nucleotide polymorphism.

[a]Number in parenthesis indicates the number of genetic factors with at least one positive result.

[b]Other variants include variable number of tandem repeats, microsatellite, duplication, insertion/deletion, SNP without rs ID, etc.

[c]Some pathways were predicted by both PBA article and *i*-GSEA4GWAS analysis.

[d]The data were from in-depth data analysis; others were from literature.

finding for each of the two (BD-SZ, BD-MDD) or three (BD-SZ-MDD) diseases. These genes were regarded as reliable cross-disorder genes, in which, *DAOA*, *DISC1*, *NRG1*, and *MTHFR* have at least three positive findings from cross-disorder studies for both BD and SZ; *SLC6A4*, *COMT*, and *MTHFR* have at least three positive findings from cross-disorder studies for both BD and MDD. Moreover, by intersection analysis of multidisease genes, BDgene also provides 127 new potential shared genes for BD-SZ, 79 for BD-MDD, and 107 for BD-SZ-MDD. Taken together, BDgene covers 412 shared genes for BD-SZ, 199 for BD-MDD, and 156 for BD-SZ-MDD. Other genetic factors shared by BD and SZ/MDD, including SNPs, haplotypes, other variants, gene–gene interactions, and chromosomal regions, are also included in BDgene. A summary of shared genetic factors between BD and SZ/MDD is shown in Table 2.

Prioritized Genes and Enriched Pathways

To rank genes by relevance to BD, we performed gene prioritization for BD candidate genes. By using 20 known BD genes as training genes, 23 prioritized genes were obtained (Table S2 in Supplement 1), in which eight genes were predicted by four tools and 15 genes were predicted by three tools. Taken together the 20 training genes and 23 prioritized genes, 43 genes were regarded as "core genes." To explore functional pathways of the core genes, pathway enrichment analysis was performed. Among the core-gene-enriched pathways, "amine binding," "synapse transmission," and "transmission of nerve impulse" were the top three statistically significant. There were 26 core-gene-enriched pathways, which have also been predicted to be associated with BD by PBA for GWAS; the top 10 of these overlapped pathways are shown in Table S3 in Supplement 1. The majority of the 10 pathways were involved in synaptic transmission, membrane, and ion channel activity. Furthermore, to provide clues for understanding disease shared etiology, we did pathway enrichment analysis for the 83 literature-reported positive genes shared by BD-SZ and 47 shared by BD-MDD; the top 10 enriched pathways for each group of shared genes are shown in Tables S4 and S5 in Supplement 1. It is worth noting that "circadian rhythm" and "rhythmic process" were two of shared pathways among the top 10 enriched pathways for both BD-SZ and BD-MDD shared genes, and several circadian rhythm–related genes such as *NPAS2*, *PER2*, *PER3*, *CRY1*, and *CLOCK* were enriched in these pathways. Specifically, circadian rhythm and rhythmic process were the top two enriched pathways for BD-MDD shared genes.

Database Usage

A powerful search engine and a detailed report for each genetic factor were provided in BDgene for researchers to search and analyze the data (Figure 2A). To facilitate researchers in browsing different types of genetic factors (SNP/gene/CNV/region) graphically and simultaneously in the context of genomic regions, gBrowse was incorporated into BDgene (Figure 2B) (56). In addition, a "My Gene Set" tool was developed to facilitate researchers in collecting genes of interest and investigating gene–gene interactions. A graphical gene network will be dynamically generated on the basis of the selected genes and their interacted genes, which might provide new candidates and clues for understanding their association with disease (Figure 2C). For example, *BDNF* and *DRD2* are two of the most studied genes in BD. When adding them into My Gene Set, the network showed that gene *CADPS2* (Ca++-dependent secretion activator 2) was interacted with both *BDNF* and *DRD2*. Although *CADPS2* has not been reported to be associated with BD, it has been shown to have high relevance to brain structure (57) and might be a potential candidate for further study. To facilitate researchers' sharing and exchange of ideas, a "Forum" module was developed. Researchers can not only make comments on specific genes or studies hosted by BDgene online in real time but also exchange ideas to review BD genetic research or upload complementary articles (Figure 2D). In summary, with the help of these tools for searching, browsing, and discussion, users can acquire an in-depth understanding of the data and the logical connections among various genetic factors with strong supporting evidence.

Discussion

As a comprehensive genetic database and the first cross-disorder database for BD, BDgene integrates multitype BD-related

Table 2. Statistics Regarding Shared Genetic Factors of BD-SZ, BD-MDD, and BD-SZ-MDD (Through March 1, 2013)

	Data Statistics		
Data Type	BD-SZ	BD-MDD	BD-SZ-MDD
Variant			
SNP	675	332	177
Haplotype	363	104	51
Others[a]	159	106	34
Gene	412	199	156
From cross-disorder studies	285	120	49
From candidate gene intersection analysis[b]	127	79	107
Gene–Gene Interaction	8	2	—
Region	40	—	—

BD, bipolar disorder; MDD, major depressive disorder; SNP, single nucleotide polymorphism; SZ, schizophrenia.

[a]Other variants include variable number of tandem repeats, microsatellite, duplication, insertion/deletion, SNP without rs ID, etc.

[b]New potential shared genes from in-depth data analysis; others were from literature.

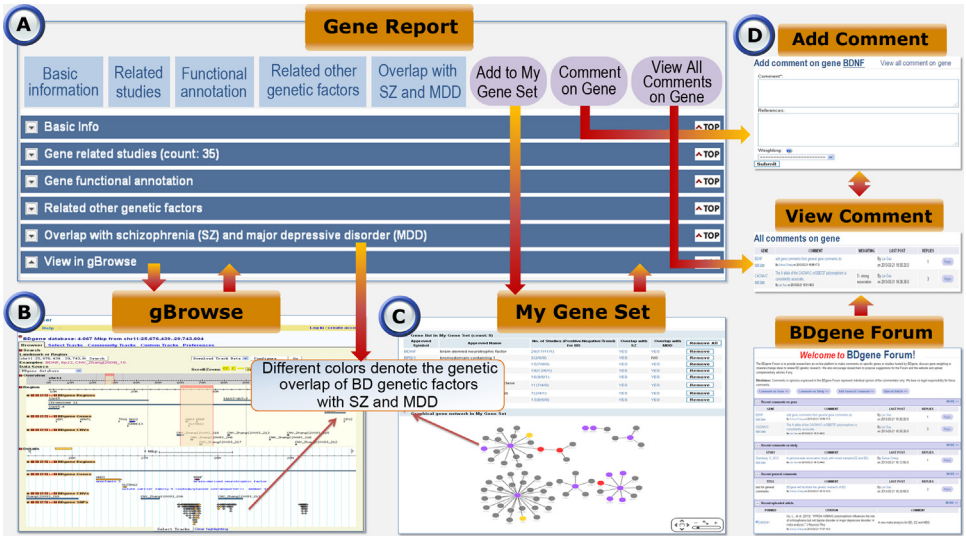

Figure 2. An overview of the connections of different tools and function modules for data access and analysis in BDgene. BD, bipolar disorder; MDD, major depressive disorder; SZ, schizophrenia.

genetic factors and factors shared with SZ and MDD through in-depth reading of 796 publications. The integration of data from the literature will provide a panoramic view of current genetic studies for BD and other highly relevant disorders. Meanwhile, the data integration makes it possible to display correlations among various genetic factors and thus to facilitate researchers to find genetic factors with multilevel evidence or multidisease evidence, which is a fundamental step in exploring valuable information for further study. By providing powerful search and browse tools, the BDgene database aims to act not only as an integrated genetic resource for BD but also as a flexible application platform for the genetic study of BD and its overlap with SZ and MDD.

On the basis of data from the literature, in-depth data mining and data analyses were conducted. Some promising results have received little attention with little additional research thus far. For example, among the 43 core genes identified to have high relevance to BD, gene *AGTR1* (angiotensin II receptor, type 1) was predicted by four tools in our gene prioritization analysis, and its variation has recently been reported to be associated with depression and frontotemporal morphology (58). However, only one candidate gene association study has noted its nominal association with BD until now (59). In addition, PBA implemented by our tool found 70 pathways, of which 57 have not been reported in published PBA articles. Among these new candidate pathways, "notch signaling pathways" has recently been reported by a network analysis combining BD GWAS data and gene expression data (60); several immune-related pathways, such as "immune effector process" and "inflammatory response," support the clinical findings that immune dysregulation and alternative inflammatory gene expression have been observed in patients with BD (61–63). Moreover, *CACNA1C* was involved in four of the top 10 pathways identified by both core-gene enrichment analysis and GWAS PBA (Table S3 in Supplement 1). This gene has been identified as associated with BD in several GWAS (37–39), and evidence from brain imaging has also demonstrated changes in total gray matter volume in carriers of risk allele (64). The pathways, in which *CACNA1C* was highly involved, might present possible explanations for this gene on its role in disease pathogenesis. In summary, the in-depth data analysis results in BDgene will provide more reliable candidates, and those that have not been intensively investigated might be worthy of further exploration.

It is worth mentioning that, in contrast to experimental research providing verification for scientific hypotheses, BDgene is data-driven research with the aim of providing more reliable candidates, as well as clues as to their possible function in explaining hypotheses of mechanism in a systematic way. Because of limited understanding of the complex mechanisms of mental disorders, the results from data-driven approaches only point to possible insights and provide hypotheses. Additional experimental replication and verification are required in future genetic, gene expression and molecular functional studies. In addition, all genetic factors included in BDgene have been collected based on studies aiming to reveal associations between candidate genetic factors and BD in diagnosed patients; therefore, the review articles without statistical results and studies focusing on endophenotypes or clinical characteristics for patients are not included in BDgene.

In conclusion, BDgene aims to fulfill the growing research needs in addressing the genetic complexity of BD and its overlap with SZ and MDD. As more research findings are published, BDgene will be updated quarterly to maintain an update-to-date resource using our established data update pipeline, which includes automatic literature search and downloading, proficient paper filtering and data extraction by research experts, and data annotation and analysis using computer programs. With the development of new research technologies, BDgene will integrate more types of data, such as that from next-generation sequencing, large-scale gene expression, and gene–environment interplay studies. Evidence from epigenetics and animal models will also be included in future efforts to update BDgene and facilitate BD genetic and mechanism research. We hope BDgene will contribute to uncovering the genetic basis of BD and related diseases, and ultimately to the improvement of disease diagnosis and treatment.

This work was supported by the State Administration of Foreign Experts Affairs of Chinese Academy of Sciences (CAS/SAFEA) International Partnership Program for Creative Research Teams (Grant No. Y2CX131003), the Knowledge Innovation Program of the Chinese

Academy of Sciences (Grant No. KSCX2-EW-J-8), and Key Laboratory of Mental Health, Institute of Psychology, Chinese Academy of Sciences. We thank Dr. Kun-Lin Zhang for his valuable suggestions on parameter selection of pathway-based analysis for genome-wide association studies.

The authors report no biomedical financial interests or potential conflict of interest.

Supplementary material cited in this article is available online at http://dx.doi.org/10.1016/j.biopsych.2013.04.016.

1. Goodwin FK, Jamison KR, Ghaemi SN (2007): *Manic-Depressive Illness: Bipolar Disorders and Recurrent Depression, 2nd ed.* New York and Oxford: Oxford University Press.
2. Merikangas KR, Akiskal HS, Angst J, Greenberg PE, Hirschfeld RM, Petukhova M, *et al.* (2007): Lifetime and 12-month prevalence of bipolar spectrum disorder in the National Comorbidity Survey replication. *Arch Gen Psychiatry* 64:543–552.
3. Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshleman S, *et al.* (1994): Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Results from the National Comorbidity Survey. *Arch Gen Psychiatry* 51:8–19.
4. Novick DM, Swartz HA, Frank E (2010): Suicide attempts in bipolar I and bipolar II disorder: A review and meta-analysis of the evidence. *Bipolar Disord* 12:1–9.
5. Hirschfeld RM, Vornik LA (2005): Bipolar disorder—costs and comorbidity. *Am J Manag Care* 11:S85–S90.
6. Barnett JH, Smoller JW (2009): The genetics of bipolar disorder. *Neuroscience* 164:331–343.
7. Huang J, Perlis RH, Lee PH, Rush AJ, Fava M, Sachs GS, *et al.* (2010): Cross-disorder genomewide analysis of schizophrenia, bipolar disorder, and depression. *Am J Psychiatry* 167:1254–1263.
8. Taylor MA, Berenbaum SA, Jampala VC, Cloninger CR (1993): Are schizophrenia and affective disorder related? Preliminary data from a family study. *Am J Psychiatry* 150:278–285.
9. Valles V, Van Os J, Guillamat R, Gutierrez B, Campillo M, Gento P, *et al.* (2000): Increased morbid risk for schizophrenia in families of inpatients with bipolar illness. *Schizophr Res* 42:83–90.
10. Smoller JW, Finn CT (2003): Family, twin, and adoption studies of bipolar disorder. *Am J Med Genet C Semin Med Genet* 123C:48–58.
11. Serretti A, Mandelli L (2008): The genetics of bipolar disorder: Genome "hot regions," genes, new potential candidates and future directions. *Mol Psychiatry* 13:742–771.
12. Offord J (2012): Genetic approaches to a better understanding of bipolar disorder. *Pharmacol Ther* 133:133–141.
13. Cordeiro Q, Talkowski ME, Chowdari KV, Wood J, Nimgaonkar V, Vallada H (2005): Association and linkage analysis of RGS4 polymorphisms with schizophrenia and bipolar disorder in Brazil. *Genes Brain Behav* 4:45–50.
14. Chowdari KV, Mirnics K, Semwal P, Wood J, Lawrence E, Bhatia T, *et al.* (2002): Association and linkage analyses of RGS4 polymorphisms in schizophrenia. *Hum Mol Genet* 11:1373–1380.
15. Li Y, Zhao Q, Zhang Z, Wang P, Che R, Tang W, *et al.* (2010): Association study between RGS4 and bipolar disorder in the Chinese Han population. *Psychiatr Genet* 20:130–132.
16. Fallin MD, Lasseter VK, Avramopoulos D, Nicodemus KK, Wolyniec PS, McGrath JA, *et al.* (2005): Bipolar I disorder and schizophrenia: A 440-single-nucleotide polymorphism screen of 64 candidate genes among Ashkenazi Jewish case-parent trios. *Am J Hum Genet* 77:918–936.
17. Prata DP, Breen G, Munro J, Sinclair M, Osborne S, Li T, *et al.* (2006): Bipolar 1 disorder is not associated with the RGS4, PRODH, COMT and GRK3 genes. *Psychiatr Genet* 16:229–230.
18. Kato T (2007): Molecular genetics of bipolar disorder and depression. *Psychiatry Clin Neurosci* 61:3–19.
19. Tranchevent L-C, Capdevila FB, Nitsch D, De Moor B, De Causmaecker P, Moreau Y (2010): A guide to web tools to prioritize candidate genes. *Brief Bioinform* 12:22–32.
20. O'Dushlaine C, Kenny E, Heron E, Donohoe G, Gill M, Morris D, *et al.* (2011): Molecular pathways involved in neuronal cell adhesion and membrane scaffolding contribute to schizophrenia and bipolar disorder susceptibility. *Mol Psychiatry* 16:286–292.
21. Manolio TA, Brooks LD, Collins FS (2008): A HapMap harvest of insights into the genetics of common disease. *J Clin Invest* 118:1590–1605.
22. Wang K, Li M, Hakonarson H (2010): Analysing biological pathways in genome-wide association studies. *Nat Rev Genet* 11:843–854.
23. Torkamani A, Topol EJ, Schork NJ (2008): Pathway analysis of seven common diseases assessed by genome-wide association. *Genomics* 92: 265–272.
24. Askland K, Read C, Moore J (2009): Pathways-based analyses of whole-genome association study data in bipolar disorder reveal genes mediating ion channel activity and synaptic neurotransmission. *Hum Genet* 125:63–79.
25. Chen L, Zhang L, Zhao Y, Xu L, Shang Y, Wang Q, *et al.* (2009): Prioritizing risk pathways: A novel association approach to searching for disease pathways fusing SNPs and pathways. *Bioinformatics* 25: 237–242.
26. Holmans P, Green EK, Pahwa JS, Ferreira MA, Purcell SM, Sklar P, *et al.* (2009): Gene ontology analysis of GWA study data sets provides insights into the biology of bipolar disorder. *Am J Hum Genet* 85:13–24.
27. Peng G, Luo L, Siu H, Zhu Y, Hu P, Hong S, *et al.* (2010): Gene and pathway-based second-wave analysis of genome-wide association studies. *Eur J Hum Genet* 18:111–117.
28. Zhang K, Cui S, Chang S, Zhang L, Wang J (2010): i-GSEA4GWAS: A web server for identification of pathways/gene sets associated with traits by applying an improved gene set enrichment analysis to genome-wide association study. *Nucleic Acids Res* 38:W90–W95.
29. Zhang L, Chang S, Li Z, Zhang K, Du Y, Ott J, *et al.* (2012): ADHDgene: A genetic database for attention deficit hyperactivity disorder. *Nucleic Acids Res* 40:D1003–D1009.
30. Lander E, Kruglyak L (1995): Genetic dissection of complex traits: Guidelines for interpreting and reporting linkage results. *Nat Genet* 11: 241–247.
31. Tranchevent L-C, Barriot R, Yu S, Van Vooren S, Van Loo P, Coessens B, *et al.* (2008): Endeavour update: A web resource for gene prioritization in multiple species. *Nucleic Acids Res* 36:W377–W384.
32. Aerts S, Lambrechts D, Maity S, Van Loo P, Coessens B, De Smet F, *et al.* (2006): Gene prioritization through genomic data fusion. *Nat Biotech* 24:537–544.
33. Chen Y, Wang W, Zhou Y, Shields R, Chanda SK, Elston RC, *et al.* (2011): In silico gene prioritization by integrating multiple data sources. *PLoS One* 6:e21137.
34. Chen J, Bardes EE, Aronow BJ, Jegga AG (2009): ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res* 37:W305–W311.
35. Chen YA, Tripathi LP, Mizuguchi K (2011): TargetMine, an integrated data warehouse for candidate gene prioritisation and target discovery. *PLoS One* 6:e17844.
36. Chang S, Zhang W, Gao L, Wang J (2012): Prioritization of candidate genes for attention deficit hyperactivity disorder by computational analysis of multiple data sources. *Protein Cell* 3:526–534.
37. Ferreira MA, O'Donovan MC, Meng YA, Jones IR, Ruderfer DM, Jones L, *et al.* (2008): Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. *Nat Genet* 40: 1056–1058.
38. Liu Y, Blackwood DH, Caesar S, de Geus EJ, Farmer A, Ferreira MA, *et al.* (2011): Meta-analysis of genome-wide association data of bipolar disorder and major depressive disorder. *Mol Psychiatry* 16:2–4.
39. Green EK, Hamshere M, Forty L, Gordon-Smith K, Fraser C, Russell E, *et al.* (2012): Replication of bipolar disorder susceptibility alleles and identification of two novel genome-wide significant associations in a new bipolar disorder case-control sample. *Mol Psychiatry*.
40. Baum AE, Akula N, Cabanero M, Cardona I, Corona W, Klemens B, *et al.* (2008): A genome-wide association study implicates diacylglycerol kinase eta (DGKH) and several other genes in the etiology of bipolar disorder. *Mol Psychiatry* 13:197–207.
41. Hattori E, Toyota T, Ishitsuka Y, Iwayama Y, Yamada K, Ujike H, *et al.* (2009): Preliminary genome-wide association study of bipolar disorder in the Japanese population. *Am J Med Genet B Neuropsychiatr Genet* 150B:1110–1117.
42. Smith EN, Bloss CS, Badner JA, Barrett T, Belmonte PL, Berrettini W, *et al.* (2009): Genome-wide association study of bipolar disorder in European American and African American individuals. *Mol Psychiatry* 14:755–763.

43. Wellcome Trust Case Control Consortium (2007): Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447:661–678.
44. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, *et al*. (2007): PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559–575.
45. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP (2011): Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 27:1739–1740.
46. Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M (2010): KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res* 38:D355–D360.
47. Flicek P, Amode MR, Barrell D, Beal K, Brent S, Chen Y, *et al*. (2011): Ensembl 2011. *Nucleic Acids Res* 39:D800–D806.
48. Allen NC, Bagade S, McQueen MB, Ioannidis JP, Kavvoura FK, Khoury MJ, *et al*. (2008): Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: The SzGene database. *Nat Genet* 40:827–834.
49. Guo L, Zhang W, Chang S, Zhang L, Ott J, Wang J (2012): MK4MDD: A Multi-level knowledge base and analysis platform for major depressive disorder. *PLoS One* 7:e46335.
50. Altshuler DM, Gibbs RA, Peltonen L, Dermitzakis E, Schaffner SF, Yu F, *et al*. (2010): Integrating common and rare genetic variation in diverse human populations. *Nature* 467:52–58.
51. Erdmann J, Grosshennig A, Braund PS, Konig IR, Hengstenberg C, Hall AS, *et al*. (2009): New susceptibility locus for coronary artery disease on chromosome 3q22.3. *Nat Genet* 41:280–282.
52. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, *et al*. (2000): Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nature Genet* 25:25–29.
53. Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, *et al*. (2009): Human Protein Reference Database—2009 update. *Nucleic Acids Res* 37:D767–D772.
54. Huang da W, Sherman BT, Lempicki RA (2009): Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protoc* 4:44–57.
55. Huang da W, Sherman BT, Lempicki RA (2009): Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 37:1–13.
56. Stein LD, Mungall C, Shu S, Caudy M, Mangone M, Day A, *et al*. (2002): The generic genome browser: A building block for a model organism system database. *Genome Res* 12:1599–1610.
57. Stein JL, Hua X, Lee S, Ho AJ, Leow AD, Toga AW, *et al*. (2010): Voxelwise genome-wide association study (vGWAS). *Neuroimage* 53: 1160–1174.
58. Taylor WD, Benjamin S, McQuoid DR, Payne ME, Krishnan RR, MacFall JR, *et al*. (2012): AGTR1 gene variation: association with depression and frontotemporal morphology. *Psychiatry Res* 202: 104–109.
59. Gratacos M, Costas J, de Cid R, Bayes M, Gonzalez JR, Baca-Garcia E, *et al*. (2009): Identification of new putative susceptibility genes for several psychiatric disorders by association analysis of regulatory and non-synonymous SNPs of 306 genes involved in neurotransmission and neurodevelopment. *Am J Med Genet B Neuropsychiatr Genet* 150B: 808–816.
60. Pedroso I, Lourdusamy A, Rietschel M, Nothen MM, Cichon S, McGuffin P, *et al*. (2012): Common genetic variants and gene-expression changes associated with bipolar disorder are over-represented in brain signaling pathway genes. *Biol Psychiatry* 72:311–317.
61. Padmos RC, Bekris L, Knijff EM, Tiemeier H, Kupka RW, Cohen D, *et al*. (2004): A high prevalence of organ-specific autoimmunity in patients with bipolar disorder. *Biol Psychiatry* 56:476–482.
62. Drexhage RC, van der Heul-Nieuwenhuijsen L, Padmos RC, van Beveren N, Cohen D, Versnel MA, *et al*. (2010): Inflammatory gene expression in monocytes of patients with schizophrenia: Overlap and difference with bipolar disorder. A study in naturalistically treated patients. *Int J Neuropsychopharmacol* 13:1369–1381.
63. Munkholm K, Vinberg M, Vedel Kessing L (2013): Cytokines in bipolar disorder: A systematic review and meta-analysis. *J Affect Disord* 144: 16–27.
64. Kempton MJ, Ruberto G, Vassos E, Tatarelli R, Girardi P, Collier D, *et al*. (2009): Effects of the CACNA1C risk allele for bipolar disorder on cerebral gray matter volume in healthy individuals. *Am J Psychiatry* 166:1413–1414.