Schizophrenia Bulletin vol. 43 no. 2 pp. 459–471, 2017
doi:10.1093/schbul/sbw102
Advance Access publication July 22, 2016

SZDB: A Database for Schizophrenia Genetic Research

Yong Wu[1,2], Yong-Gang Yao[1–4], and Xiong-Jian Luo[*,1,2,4]

[1]Key Laboratory of Animal Models and Human Disease Mechanisms of the Chinese Academy of Sciences and Yunnan Province, Kunming Institute of Zoology, Kunming, China; [2]Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming, China; [3]CAS Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Shanghai, China

[4]YGY and XJL are co-corresponding authors who jointly directed this work.

*To whom correspondence should be addressed; Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan 650223, China; tel: +86-871-68125413, fax: +86-871-68125413, e-mail: luoxiongjian@mail.kiz.ac.cn

Schizophrenia (SZ) is a debilitating brain disorder with a complex genetic architecture. Genetic studies, especially recent genome-wide association studies (GWAS), have identified multiple variants (loci) conferring risk to SZ. However, how to efficiently extract meaningful biological information from bulk genetic findings of SZ remains a major challenge. There is a pressing need to integrate multiple layers of data from various sources, eg, genetic findings from GWAS, copy number variations (CNVs), association and linkage studies, gene expression, protein–protein interaction (PPI), co-expression, expression quantitative trait loci (eQTL), and Encyclopedia of DNA Elements (ENCODE) data, to provide a comprehensive resource to facilitate the translation of genetic findings into SZ molecular diagnosis and mechanism study. Here we developed the SZDB database (http://www.szdb.org/), a comprehensive resource for SZ research. SZ genetic data, gene expression data, network-based data, brain eQTL data, and SNP function annotation information were systematically extracted, curated and deposited in SZDB. In-depth analyses and systematic integration were performed to identify top prioritized SZ genes and enriched pathways. Multiple types of data from various layers of SZ research were systematically integrated and deposited in SZDB. In-depth data analyses and integration identified top prioritized SZ genes and enriched pathways. We further showed that genes implicated in SZ are highly co-expressed in human brain and proteins encoded by the prioritized SZ risk genes are significantly interacted. The user-friendly SZDB provides high-confidence candidate variants and genes for further functional characterization. More important, SZDB provides convenient online tools for data search and browse, data integration, and customized data analyses.

Key words: schizophrenia/genetic study/database/gene expression/integrative analysis

Introduction

Schizophrenia (SZ) is a severe mental disorder characterized by abnormal perceptions, incoherent or illogical thoughts, and disorganized speech and behavior. It affects approximately 0.5%–1% of the world populations[1] and is one of the leading causes of disability worldwide.[2–4] With the substantial morbidity and high mortality, SZ imposes profound impact on patients' quality of life, and significant economic burden on families and society.[4] It is estimated that the overall cost of SZ in the United States in 2002 was $62.7 billion.[5] In Canada, the total cost of SZ was estimated to be 6.85 billion (Canadian dollars) in 2004.[6] The global economic burden of SZ is continuously increasing, which makes SZ an urgent global health issue.

Despite the fact that millions of people are suffering from SZ and the substantial personal and societal costs, current available antipsychotic medications can only alleviate the symptoms of SZ. Moreover, the efficacy of antipsychotic drugs varies greatly among patients (ie, many patients are treatment-resistant).[7,8] So far, there is no drug or treatment that can completely cure SZ or ensure that there will be no further psychotic episodes. A key reason for this therapeutic conundrum is due to the unknown pathophysiological mechanism of SZ. During the past decades, great efforts have been made to identify the causes of SZ and to develop new drugs and treatments. Though significant progress has been achieved, the etiology of SZ remains largely unknown.

A growing body of evidence strongly suggest that both genetic and environmental factors are involved in the pathophysiology of SZ.[9] Previous studies have shown that environmental factors such as obstetric complications, exposure to influenza viruses during gestation, and prenatal exposure to maternal stress can increase risk of SZ.[10–15] Though environmental factors have a role in the pathogenesis of SZ, high heritability clearly shows the

major role of inherited genetic variants in the etiology of SZ.[16] Family, twin and adoption studies have consistently demonstrated the key role of genetic components in the pathogenesis of SZ. The heritability estimate of SZ is about 0.80,[16] which is higher than most of other complex diseases.[17]

To elucidate the genetic susceptibility to SZ, large numbers of linkage and association studies, as well as meta-analyses have been performed, and numerous risk variants, genes, and chromosomal regions have been reported to be associated with SZ.[18–20] Despite the fact that many promising candidate genes have been identified, the sample sizes are relatively small and the coverage of genetic markers are relatively low in these studies, which limited the identification of credible susceptibility variants and genes. The advent of genome-wide association studies (GWAS) provides critical opportunities to explore the genetic landscape of SZ. In 2008, *ZNF804A* had been identified as the first gene that reached a genome-wide significant level for SZ.[21] Since then, multiple GWAS have been conducted in world populations and numerous SZ-associated variants and loci were identified.[22–35] A landmark of SZ genetics is the establishment of Psychiatric Genomics Consortium (PGC), which performed the largest GWAS meta-analysis of SZ so far and identified numerous high-confidence genetic risk variants and loci.[36]

With the continuous increase of sample size in PGC, novel risk variants and genes can now be identified at unprecedented rates. Despite the great success of GWAS in identifying SZ risk variants, understanding the pathological role of these risk variants in SZ remains a daunting task. First, most of the SZ-associated variants identified by GWAS reside in noncoding regions (with limited annotation and no obvious functional consequence), how these risk variants contribute to SZ risk is largely unknown. There is a growing gap between the identification of SZ risk variants and the extraction of meaningful biological information. Second, besides the common genetic variants (single nucleotide polymorphisms, SNPs) identified by GWAS, other types of genetic variations including rare variants, copy number variation (CNV), and structural genomic variations play a pivotal role in SZ.[37–41] Therefore, there is a pressing need to integrate data from various genetic studies of SZ to provide a global view of current genetic findings on SZ. Third, systematic integration and in-depth analysis of multi-type data from various sources (eg, genetic and gene expression studies) will help to provide useful guide for further experimental validation and to establish the connection among different types of data. For example, selecting high-confidence candidate gene for further functional validation is challenging due to a large number of genes identified by genetic association studies. Gene prioritization analysis (eg, convergent functional genomics)[42] provides a valuable solution for selecting the candidate genes for experimental verification. Fourth, accumulating evidence imply that common polygenic variation contributes to SZ risk.[22] That is, SZ is likely caused by multiple genes that disrupt one or more molecular pathways.[43,44] Thus, pathway and network-based analysis is an effective method to identify the molecular pathways dysregulated in SZ and helps with mechanism investigation and drug development. To better understand the genetic architecture of SZ and to facilitate the translation of genetic findings into molecular risk mechanisms for SZ, we developed SZDB, a comprehensive database containing multiple integrative data from various layers of SZ genetic researches. SZDB not only provides a user-friendly web interface for data search, browse and visualize, but also provides in-depth data integration and analysis functions such as SNP functional annotation, spatio-temporal gene expression pattern analysis in human brain, expression quantitative trait loci (eQTL) and network-based (PPI and co-expression) analyses.

Methods

SNP Associations From GWAS of SZ

Several GWAS of SZ have been performed during the past decade. To further increase the statistic power, PGC2 performed the largest meta-analysis of SZ GWAS so far.[36] In brief, genome-wide genotype data of 49 ancestry-matched, nonoverlapping case-control samples and 3 family-based samples of European ancestry were obtained. The total sample size included in the GWAS meta-analysis was 82 315, including 35 476 SZ cases and 46 839 controls. After quality control and SNP imputation, association test was performed separately in each sample and the results were combined and meta-analyzed. More detailed information about sample description, genotyping and statistical analyses can be found in the original paper.[36] We downloaded the summary statistics (SNP associations) of PGC2 from the PGC website (http://www.med.unc.edu). In total, summary statistics of 9 444 230 SNPs were downloaded.

Plotting of Regional Association Signal

To visualize the results of genome-wide association data of PGC2,[36] we utilized the LocusZoom[45] to plot regional association results. The *P* values of SNPs were downloaded from the PGC website (http://www.med.unc.edu/pgc/downloads).[36] We downloaded the Hg19 genome assembly and 1000 Genomes population data from http://locuszoom.sph.umich.edu/locuszoom/.

SNP Function Annotation

Most of the SZ-associated SNPs identified by GWAS lie in noncoding region and are predominately located within putative regulatory elements,[36] suggesting that these variants may confer risk to SZ through affecting

gene expression. We took advantage of the recent study by Boyle et al[46] that systematically annotated SNPs in human genome using experimental data sets from ENCODE[47] and other sources.[46,48] Briefly, high-throughput computational and experimental data sets, including DNase footprinting, binding motif, ChIP-seq, chromatin state, eQTL, as well as other sources were used to predict if a specific SNP has a regulatory potential. To make the results easier to use and understand, a score system was introduced and a corresponding database (RegulomeDB, http://RegulomeDB.org/) was developed by Boyle and coworkers.[46] Lower scores in the RegulomeDB suggest increasing evidence for a variant to be located in a functional region. More detailed information about variant annotation and RegulomeDB database can be found in the original paper.[46] We downloaded the RegulomeDB score data for all SNPs (including genotyped and imputed) analyzed in PGC2[36] and integrated these scores into SZDB. We also annotated the potential functional consequences of the non-synonymous (missense) variants using SIFT[49,50] and PolyPhen-2[51], 2 extensively used prediction tools to conduct in silico functional prediction of non-synonymous (missense) variants.

Extraction and Compilation of SZ Candidate Genes

Previous studies (including genetic linkage and association studies,[18,19] GWAS,[36] integrative studies such as convergent functional genomics (CFG)[42] and *Sherlock* integrative analysis,[52] and whole exome sequencing studies[53,54]) have identified numerous SZ candidate genes. We systematically extracted and compiled a comprehensive list of SZ candidate genes from these studies. Based on the study type, SZ candidate genes were classified into the following categories:

(1) Genes identified by genetic linkage and association studies. To systematically combine the results of previous genetic linkage studies, Lewis et al[18] conducted a rank-based genome scan meta-analysis and identified multiple chromosomal regions that may harbor SZ susceptibility variants and genes. In total, 160 genes located in these chromosomal regions were extracted (supplementary table S1). Recently, Ng et al[19] also performed a systematic meta-analysis using 32 genome-wide linkage studies of SZ. We extracted all 173 genes identified by Ng et al[19] (supplementary table S2).

In addition to linkage studies, numerous SZ candidate genes have been identified by association studies and we included these genes in our list. Briefly, we extracted 42 top genes from the SZGene database (http://www.szgene.org/) prioritized by Allen et al,[20] who performed a systematic meta-analysis and identified multiple SZ risk genes (supplementary table S3). We also extracted SZ candidate genes from the SZGR database (formerly available at http://bioinfo.mc.vanderbilt.edu/SZGR/, but now could not be accessible) developed by Jia et al.[55] Through using combined odds ratio (COR) method, Sun et al[56] ranked and prioritized 38 core (supplementary table S4) and 75 SZ candidate genes (supplementary table S5) and these genes were included in the SZGR database.[55] We downloaded and integrated all these SZ candidate genes into SZDB, and a total of 226 nonoverlapping SZ candidate genes (supplementary table S6) were included.

(2) Genes identified by GWAS of SZ. Multiple GWAS of SZ had reported numerous risk genes.[22–35] We manually extracted SZ risk genes that reached the genome-wide significance level. Totally, 375 nonoverlapping genes (supplementary table S7) were extracted and integrated into SZDB.

(3) Genes identified by convergent functional genomics (CFG). By using a convergent functional genomics approach, Ayalew et al[42] prioritized 42 promising SZ candidate genes (supplementary table S8) and we included these genes into SZDB.

(4) Genes affected by CNVs. Accumulating evidence suggests that CNVs play an important role in SZ. We systematically prioritized the genes affected by CNVs in SZ and identified pivotal genes that are frequently disrupted by CNVs in SZ cases.[57] We extracted these genes (supplementary table S9) and integrated them into SZDB.

(5) Genes identified by *Sherlock* integrative analysis.[58] Multiple lines of evidence support the important role of gene expression dysregulation in SZ. Recently, He et al[58] developed a statistical framework (named *Sherlock*) to identify disease-associated genes through matching eQTL and SNP associations from GWAS. We performed integrative analysis using *Sherlock* and identified 12 promising SZ candidate genes.[52] We integrated these 12 genes into SZDB.

(6) Genes differentially expressed in SZ cases. Dysregulation of gene expression has a key role in SZ pathogenesis. To detect genes that are differentially expressed between SZ patients and healthy controls, many large-scale gene expression studies have been performed. We downloaded 5 microarray datasets (which used brain tissues from SZ patients and controls) from the gene expression omnibus (GEO; [http://www.ncbi.nlm.nih.gov/geo/]), including GSE53987 (contains 3 brain tissues: prefrontal cortex, striatum and hippocampus; a total of 114 samples),[59] GSE12649 (prefrontal cortex; a total of 69 postmortem brains samples),[60] GSE21138 (prefrontal cortex; a total of 59 postmortem brains samples),[61] GSE35978 (cerebellum and parietal cortex brain; a total of 195 samples),[62] and GSE62191 (frontal cortex; a total of 59 samples).[63] We re-analyzed these microarray datasets, and top 1% differentially expressed genes in SZ cases and controls were integrated into SZDB. The differentially expressed genes identified by RNA sequencing (RNA-seq)[64] were also integrated into SZDB.

(7) Genes identified by *Pascal* gene-based test. The gene-based analysis is a strategy to improve statistical power and gain biological insight.[65] Through this test, one can get an overview from the gene and pathway level rather than the SNP level. Here we used the gene-based test Pathway scoring algorithm (*Pascal*)[65] to integrate PGC2 GWAS data. A total of 343 genes with *P*-values lower than 10^{-6} were included in the SZDB (supplementary table S10).

(8) Genes identified by whole exome sequencing. Accumulating evidence suggest that *de novo* mutations and rare disruptive mutations play an important role in SZ pathogenesis. Recently, Fromer et al[54] performed the largest exome sequencing study of *de novo* mutations in SZ and they found that *de novo* mutations in SZ are overrepresented in synaptic networks. We integrated genes (supplementary table S11) that carrying *de novo* mutations (including coding and canonical splice site variants) in SZ probands from the study of Fromer et al[54] into SZDB. In addition, Purcell et al[53] identified numerous genes that contain rare disruptive mutations in SZ cases through exome sequencing. We also integrated the genes (supplementary table S12) identified by Purcell et al[53] into SZDB.

(9) Genes differentially methylated in SZ. In addition to genetic components, environmental factors may also play an important role in SZ. DNA methylation is a common epigenetic modification that is influenced by environmental factors, and recent studies[66,67] have identified multiple differentially methylated genomic regions (or genes) among SZ cases and healthy controls. We integrated the differentially methylated genes identified by Montano et al[67] and Aberg et al[66] into SZDB. In addition, we also integrated the methylation quantitative trait loci from study of Hannon et al[68] into SZDB.

In-depth Data Analysis and Systematic Integration

Based on the above comprehensive data/ gene lists from various sources, we carried out the following in-depth data analyses and systematic integration:

(1) eQTL analysis. Genetic variants affecting gene expression in brain tissues may confer risk of SZ. To investigate if genome-wide significant SNPs identified by PGC2 are associated with gene expression in human brain, we downloaded genome-wide expression and genotype data from the study of Myers et al,[69] which included 193 neuropathologically normal human brain samples. The association between each SNP and transcript was assessed by using PLINK,[70] as described in the study of Myers et al.[69] In total, we assessed correlations among 366 140 SNPs and the expression of 14 078 transcripts. Briefly, before the analysis, rank-invariant normalized expression data were $\log_{10}$ transformed, and missing data were encoded as missing, not as a zero level of expression. We excluded transcripts that were expressed in less than 5% of the series from the analysis. The following minimum SNP cut-off values were used: per sample call rate at least 90%, per SNP call rate at least 90%, per SNP minor allele frequency of at least 1%, and lack of significance ($P > .05$) for the Hardy-Weinberg equilibrium tests. Only those pairs with transcript-specific empirical *P* values ≤ .05 were retained. SNP position and gene position were based on the Hg19 genome assembly.

(2) Spatio-temporal expression pattern analysis. A pivotal step to elucidate the role of risk genes in SZ pathogenesis is to understand the spatio-temporal expression pattern of risk genes in brain. To investigate if SZ risk genes are preferentially expressed in specific brain tissues (eg, prefrontal cortex and hippocampus) and at specific developmental stages (eg, fetal and postnatal), we performed spatio-temporal gene expression pattern analysis using the RNA-seq based gene expression data from the BrainSpan: Atlas of the Developing Human Brain (http://www.brainspan.org).[71] These expression data were from 16 brain regions (a total of 42 postmortem brains) that spanned 15 consecutive periods of neurodevelopment and adulthood from 8 post-conceptual weeks (PCW) to 40 years. The data were partitioned into different developmental stages and subsets of brain regions as described by Willsey et al.[72] Data processing, quality control and statistical analyses were performed as described previously.[73] In addition to expression data from the BrainSpan,[71] we also downloaded expression data from the BrainCloud (http://braincloud.jhmi.edu/),[74] which contains gene expression data from prefrontal cortex. The developmental stages were also partitioned into 15 periods. More detailed information about the BrainCloud can be found in the original study[74] and the SZDB website (http://www.szdb.org/index.html).

(3) Co-expression analysis. Recent study suggested that SZ risk genes are significantly co-expressed in patterns specific to developmental stage and neuroanatomical structure.[73] Therefore, co-expression analysis may help to prioritize SZ candidate genes from large-scale genetic studies (eg, GWAS or transcriptome). To explore if SZ risk genes were co-expressed in specific brain regions at specific developmental stages, we performed gene co-expression analyses using the RNA-seq based expression data from the BrainSpan: Atlas of the Developing Human Brain (http://www.brainspan.org/).[71] Briefly, we divided the tissues from the BrainSpan atlas into 4 anatomic regions and 15 developmental stages as described by Willsey et al[72] and Gilman et al.[75] The 4 anatomic regions are: (1) V1C-STC cluster (V1C, ITC, IPC, A1C, and STC); (2) Prefrontal and primary motor-somatosensory

cortex or M1C-S1C cluster (M1C, S1C, VFC, MFC, DFC, and OFC); (3) STR-AMY cluster (STR, HIP, AMY); (4) MD-CBC cluster (MD and CBC). The 15 developmental stages are: (1) Embryonic, from 4 to 8 post-conception weeks (PCW); (2) Early fetal, from 8 to 10 PCW; (3) Early fetal, from 10 to 13 PCW; (4) Early mid-fetal, from 13 to 16 PCW; (5) Early mid-fetal, from 16 to 19 PCW; (6) Late mid-fetal, from 19 to 24 PCW; (7) Late fetal, from 24 to 38 PCW; (8) Neonatal & early infancy, from 0 to 6 months (M); (9) Late infancy from 6 to 12 M; (10) Early childhood, from 1 to 6 years (Y); (11) Middle and late childhood, from 6 to 12 Y; (12) Adolescence, from 12 to 20 Y; (13) Young adulthood, from 20 to 40 Y; (14) Middle adulthood, from 40 to 60 Y; (15) Late adulthood, more than 60 years old. Pearson correlation coefficients were calculated in each brain region across developmental stages as described previously.[72,73] Only gene pair with a Pearson correlation coefficient equal or bigger than 0.8 was retained.

(4) Protein–protein interaction (PPI) analysis. Accumulating evidence suggest that proteins involved in the same disease are more likely to interact with each other (ie, guilt-by-association).[76,77] We had showed that SZ risk genes encode an highly interconnected network,[78,79] implying PPI information may help to prioritize SZ candidate genes. To explore if proteins encoded by SZ candidate genes are physically interacted with proteins encoded by other genes (including SZ risk genes and others), we downloaded PPI data from InWeb, a well-characterized PPI database developed by Lage et al.[80,81] Genes directly interact with SZ genes were regarded as the first-degree interaction genes.

(5) Analysis of differentially expressed genes. Genes differentially expressed in SZ cases and healthy controls may play a pivotal role in the etiology of SZ. Though multiple gene expression studies have been performed to identify differentially expressed genes in SZ,[59–63] we noticed that only limited overlapping genes were identified among different studies. To compare the differentially expressed from different studies, we preformed differentially expression analyses using microarray datasets from GEO. The main inclusion criteria are as follows: First, GEO datasets with a relatively large sample size (more than 20 SZ cases) were selected. Second, GEO datasets using brain tissues (including prefrontal cortex, hippocampus, and striatum) were given higher priority. Third, GEO datasets that contain genome-wide gene expression data were selected. It should be noted that only limited expression datasets are public available and can be downloaded from GEO. Based on these criteria, we selected 5 GEO datasets[59–63] for differential expression analysis. GEO2R (http://www.ncbi.nlm.nih.gov/geo/geo2r/), which uses limma (Linear Models for Microarray Analysis) in R package (an Empirical Bayes method), was used to analyze differentially expressed genes. GEO2R includes the following steps: (1) Load series and platform data from GEO; (2) Group names for all samples; (3) Transform expression data by log2 (if the data were not transformed, auto-detect by GEO2R); (4) Set up the data and proceed with analysis; (5) Load NCBI platform annotation; (6) Replace original platform annotation. To visualize these results, we used the jQuery plugin plotly (https://plot.ly/javascript/) to draw a boxplot for every gene in SZ cases and controls. We also downloaded the differentially expressed genes (a total of 798 genes) identified by the RNA-seq[64] and integrated these results into SZDB.

(6) Co-expression module analysis. To explore if SZ risk genes show similar expression pattern during cortical development, we performed co-expression network analysis using the RNA-seq based expression data from the BrainSpan.[71] Signed weighted gene co-expression network analysis was performed as previously described.[82]

(7) Gene-based analysis. Currently, most of the GWAS tested the association between a trait and individual genetic variant (usually SNP). Gene-based method is a useful complement to GWAS as it considers association between a trait and all variants (eg, SNPs) within a gene rather than each marker individually. To identify the genes that are associated with SZ at the gene level, we performed gene-based analysis using the *Pascal* software (summary statistics, ie, SNP *P* values from PGC2 were used as input).[65] More detailed information can be found in the supplementary material.

Prioritization of SZ Risk Genes

We used an arbitrary cumulative scoring method to prioritize SZ risk genes. The rationale of our method assumed that SZ-associated gene may be detected in several independent studies (or datasets). For example, if a specific gene was identified to be associated with SZ in genetic studies and gene expression study also showed that this gene is dysregulated in SZ cases, we would score this gene as a promising SZ candidate gene. More detailed information about this polyevidence scoring approach can be found in the paper of Ayalew et al[42] in which they identified promising SZ candidate genes through integrating multiple lines of evidence from different SZ researches. We took advantage of the following lines of evidence to prioritize SZ candidate genes in SZDB:

(1) Genes identified by GWAS. If a gene was reported to be associated with SZ in GWAS, the total score of this gene rises by 1 point.

(2) Genes identified by genetic linkage and association studies. If a gene was reported to be associated with SZ in previous genetic linkage or association studies, the total score of this gene rises by 1 point.

(3) Genes identified by convergent functional genomics (CFG). If a gene was in the top gene list (supplementary table S8) prioritized by CFG,[42] the total score of this gene rises by 1 point.
(4) Genes frequently disrupted by CNVs. If a gene was frequently disturbed by CNVs in SZ (supplementary table S9),[57] the total score of this gene rises by 1 point.
(5) *Sherlock* integrative analysis. If a gene was identified to be associated with SZ by *Sherlock* integrative analysis,[52] the total score of this gene rises by 1 point.
(6) Genes differentially expressed in SZ cases and controls. If the expression level of a specific gene was reported to be significantly changed in SZ cases compared with controls, the total score of this gene rises by 1 point.
(7) Gene-based evidence. If a gene showed a significant association with SZ in the gene-based test, the total score of this gene rises by 1 point.

We summed up all scores for each gene and those genes with a score ≥2 were regarded as the prioritized genes (supplementary table S13).

Drug Target Identification

We explored whether the top 29 prioritized genes in SZDB are drug targets through using the Drug-Gene interaction database (http://dgidb.genome.wustl.edu/)[83] and the list of 1030 druggable genes compiled by Rask-Andersen et al.[83] The druggable gene list by Rask-Andersen et al[83] was developed by systematically integration of public available druggable targets resources, including the DrugBank database (http://www.drugbank.ca),[84] the Drugs@FDA database (http://www.accessdata.fda.gov/scripts/cder/drugsatfda/), and the Drugs in Clinical Trials Database (http://www.centerwatch.com).

Results

Database Overview

On the basis of comprehensive data collection and curation, we performed in-depth data analysis and systematic integration and included all these results in the SZDB. Therefore, SZDB not only contains genetic risk variants and susceptibility genes for SZ, but also provides a demonstration of the in-depth analytical results, such as SNP function annotation, eQTL, spatio-temporal expression and co-expression, and PPI data. Specifically, SZDB integrates powerful online analysis tools and users can easily perform custom analyses such as eQTL, PPI, co-expression and spatio-temporal expression pattern analyses (figure 1). All of the preprocessed datasets can be freely downloaded at SZDB (http://www.szdb.org/download.html). The user-friendly web interface of the SZDB contains 3 core modules: Search module, Analysis and Tools module, and Genes module.

(1) Search module. This module provides a powerful search engine for SNP and gene query. Users can input interested SNPs and genes in the search bar. Multiple search items (eg, a set of interested SNPs or genes) are allowed for single query. For SNP query, SZDB will generate a detailed report (supplementary figure S1), including chromosomal location based on the Hg19 genome assembly and corresponding genomic coordinates of the query SNP, SNP alleles (polymorphism), SNP location in gene (flanking or nearby genes of the query SNP), SNP type (eg, synonymous or non-synonymous substitution), association significance (OR and *P* value) with SZ (from PGC2[36]), eQTL (genes whose expression are associated with the query SNP),[69] regulatory potential (RegulomeDB score[46,48]). If the query SNP is a non-synonymous SNP, SZDB will also predict the potential impact of amino acid change on protein structure and function using SIFT[49,50] and PolyPhen-2.[51] To visualize the association significance between the query SNP and SZ, a LocusZoom link for the query SNP is provided (supplementary figure S1). The LocusZoom generates and plots the regional association results between the query SNP (including SNPs that are located 200 kb up and downstream of the query SNP) and SZ using GWAS data from PGC2.[36] The query SNP is highlighted in the LocusZoom results, so the users can evaluate if the query SNP has the smallest association significance (*P* value) with SZ. For gene query, SZDB will return a "Results" page which includes the Entrez gene id, genomic coordinate, cytogenetic location, and brief description of the query gene (supplementary figure S2). Of note, a LocusZoom link for the query gene is also provided, and the users can download and browse the association significance between SNPs located in this gene and SZ (based on PGC2 data[36]; supplementary figure S3). Through using the LocusZoom, the users can get an overview about the association between the query gene and SZ (based on PGC2 data[36]). The most significant SNP is highlighted in the LocusZoom results (supplementary figure S3). Moreover, SZDB provides the differentially expressed genes in SZ cases and healthy controls.
(2) Analysis and Tools module. This is the most important and useful module. Results from our in-depth analyses were integrated by this module. In this module, users can perform customized analyses, including co-expression analysis, spatio-temporal expression pattern analysis, protein-protein interaction analysis, association analysis, and eQTL analysis. Detailed descriptions for each analysis are as follows: (a) Co-expression analysis. The co-expression analysis can provide exploratory data for candidate gene prioritization and suggestion for more work (eg, functional experiments) to validate if the candidate gene

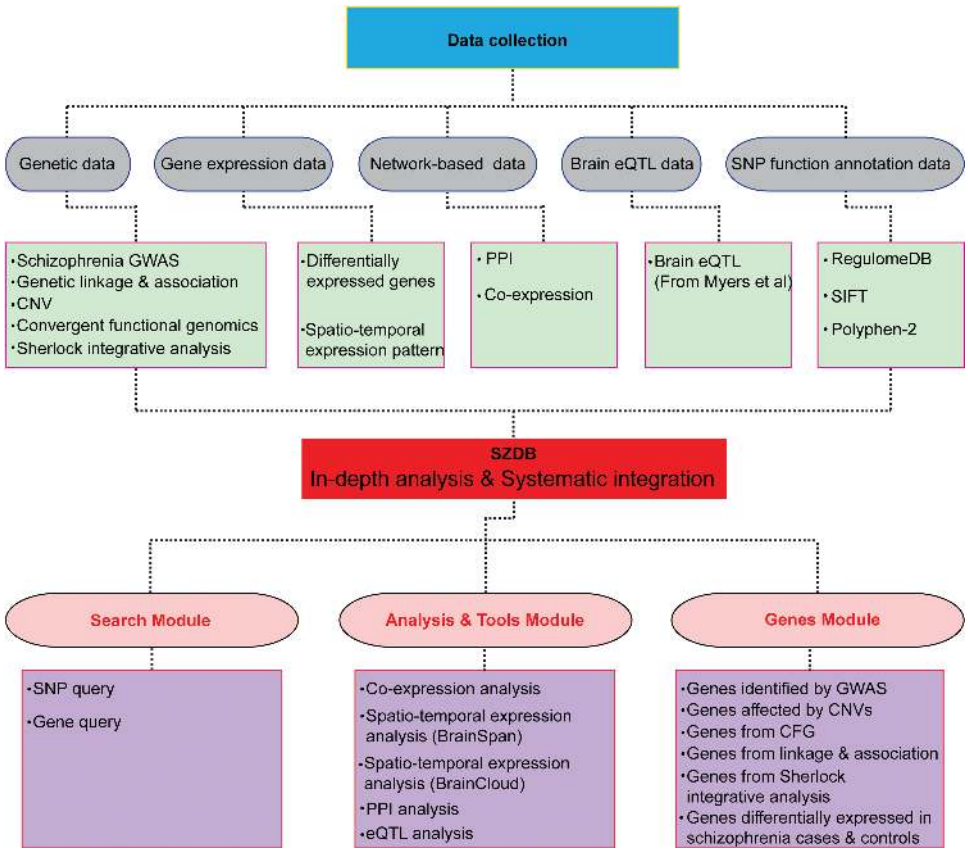

Fig. 1. Overview of SZDB database. Multi-type data from various sources were collected and curated. On the basis of comprehensive data collection and curation, in depth-analyses and systematic integration were performed. SZDB contains 3 core modules, which can provide comprehensive description information and perform customized analysis.

contributes to SZ pathogenesis. Users can explore if a specific gene (or multiple genes) is significantly co-expressed with other genes in a specific brain region at a specific developmental stage (supplementary figure S4). For each input gene, SZDB will return the genes that have the most significant co-expression correlation coefficient (default Pearson correlation coefficient ≥ 0.8) with the input gene. (b) Spatio-temporal expression pattern analysis. Two well-characterized expression datasets[74,85] were used to construct the spatio-temporal expression pattern of SZ risk genes. Multiple genes are allowed as input for a single query. For each query, SZDB generates the spatio-temporal plot of the query genes. The average expression level of all genes is also calculated and displayed. As the expression data from the BrainCloud[74] contains only one brain tissue (prefrontal cortex), the default mode will calculate the temporal expression patterns of the SZ risk genes in prefrontal cortex. For the Brainspan,[71] users can select different brain regions. Users can conduct customized analysis to examine the spario-temporal expression pattern of the query genes (supplementary figure S5). (c) PPI analysis. For a specific input protein (or proteins), SZDB extracts all of the proteins that interacted with the query protein (supplementary figure S6). Multiple proteins are allowed as input for a single query. For visualization, SZDB draws the PPI plot based on the interactions among the input proteins. (d) Regional association plot and linkage disequilibrium analysis. To visualize the association results, SZDB generates regional association plot (supplementary figure S3). Users can extract the associations (*P* values) between genetic variants located in a specific chromosomal region and SZ using the PGC2 GWAS data.[36] Three types of query, including SNP, gene, and region are allowed. Users first input the interested gene (SNP or region), then select the flanking region (from several to hundreds kb). SZDB generates the regional association plot and highlights the most significant SNP in the queried region. (e) eQTL analysis. We integrated brain eQTL data from Myers et al[69] into SZDB and users can perform SNP or gene query. For SNP query, SZDB will carry out the eQTL analysis to identify the gene (or genes) whose expression is associated with the query SNP. For gene query, SZDB will extract the SNP (or SNPs) that is associated with the expression of the query gene (supplementary figure S7).

(3) Genes module. Multiple SZ candidate genes have been identified by different methods, including genetic

linkage and association studies, differential gene expression studies, convergent functional genomics, *Sherlock* integrative analysis and so on. Currently, SZ candidate genes from 9 resources were collated and integrated into SZDB: (a) Genes frequently affected by CNVs in SZ cases; (b) Genes identified by SZ GWAS; (c) Genes identified by genetic linkage and association studies; (d) Genes identified by convergent functional genomics (CFG); (e) Genes identified by Sherlock integrative analysis; and (f) Genes differentially expressed in SZ cases and controls; (g) Genes identified by the gene-based test (*Pascal*)[65]; (h) Genes identified by whole exome sequencing; and (i) Genes differentially methylated in SZ cases compared with controls. Users can perform customized queries to retrieve these genes.

Prioritized SZ Genes, Enriched Pathways and Drug Targets

In addition to providing a user-friendly online resources and powerful search and analysis tools for users, we also ranked SZ risk genes through using a cumulative scoring strategy (see methods). Polyevidence scoring identified 29 top prioritized SZ candidate genes that have a final score of 3 and above (figure 2 and supplementary table S13). Among these top prioritized genes, 3 (*DRD2*, *PGBD1* and *DOC2A*) have the highest final score (4 points, ie, 4 lines of evidence support these genes may be associated with SZ), suggesting that these 3 genes may represent promising SZ risk genes. We performed the pathway analysis[86,87] using genes that have a polyevidence score of 2 and above and identified significant enriched pathways, including synaptic transmission, dopamine metabolic process and ion channel activity (table 1).

To explore if the top prioritized genes in SZDB are drug targets, we examined the Drug-Gene interaction database[83] and the druggable gene list of Rask-Andersen et al.[83] We found that 7 of the top 29 genes prioritized in SZDB database are drug targets, including *DRD2*, *BDNF*, *COMT*, *CNR1*, *GRIA1*, *SLC1A2*, and *SRR* (supplementary table S14). Of note, *DRD2* is the most intensively studied gene and is the target of most of drugs.

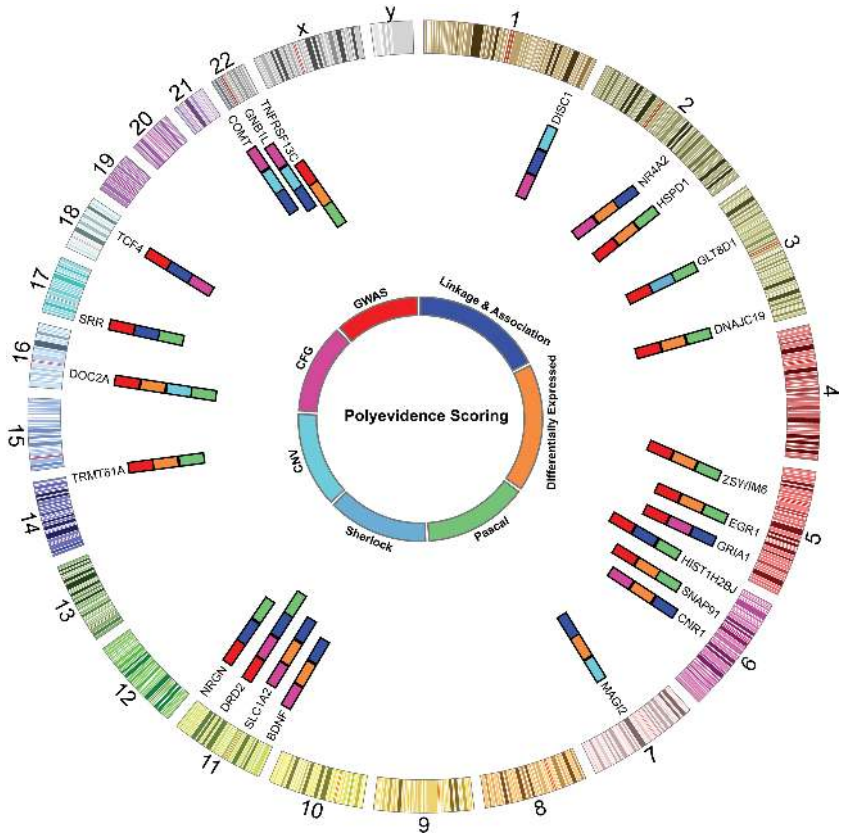

Fig. 2. Polyevidence scoring strategy and distribution of top prioritized schizophrenia risk genes on human genome. Schizophrenia risk genes were collected from 7 different sources, including genetic study (genome-wide association studies [GWAS], linkage and association, copy number variations [CNVs]), gene-based analysis, gene expression study, convergent functional genomics, *sherlock* integrative analysis, and differentially expressed genes in schizophrenia cases and controls. Evidence from these sources was integrated and genes were scored based on their presence in these sources.

Genes Implicated in SZ are Highly Co-expressed in Human Brain

On the basis of whole-genome transcriptomic data from the BrainSpan (http://www.brainspan.org),[71] we constructed gene co-expression networks by applying signed weighted gene co-expression network analysis (supplementary methods). In total, we identified 23 co-expression modules (figure 3). To further explore if SZ candidate genes are enriched in specific co-expression modules, we used 2 datasets of SZ candidate genes. The first dataset included the genes identified by recent GWAS of SZ (supplementary table S7). The second dataset included genes from our prioritized SZ candidate genes (ie, genes with a ployevidence score of 2 and above; supplementary table S13). We found that SZ candidate genes from GWAS are significantly enriched in co-expression module 2 (M2) (P = .0083) and M5 (P = .0079; figure 3). Intriguingly, genes from our polyevidence scoring were also significantly enriched in M2 (P = .018) and M5 (P = .019). These results suggested that SZ candidate genes are highly co-expressed in human brain.

Proteins Encoded by Prioritized SZ Risk Genes are Significantly Interacted

We performed PPI analysis (detailed analysis procedure can be found in our previous study[78]) using the prioritized SZ genes from SZDB (the genes that have a final score of 2 and above; supplementary table S13). We found that proteins encoded by prioritized SZ risk genes are significantly interacted (P < .05) and formed a densely

Table 1. Gene Ontology (GO) Analysis of Genes That Have a Polyevidence Score of 2 and Above

Category	Term	Genes	P_{adj}
GOTERM_BP_FAT	Synaptic transmission	*HTR2A, BDNF, CHRNA3, CHRNB4, DRD1, DRD2, DRD4, EGR1, GRIA4, GRIN2B, GRM3, NISCH, SLC1A3, ERBB4*	5.60×10^{-5}
GOTERM_BP_FAT	Dopamine metabolic process	*COMT, DRD1, DRD2, DRD4, MAOA, NR4A2, SLC6A3*	1.20×10^{-4}
GOTERM_BP_FAT	Ion channel activity	*CHRNA7, CACNA1I, CACNB2,CHRNA3, CHRNA5, CHRNB4, GABRB3, GRIA4, GRIA1, GRIN2B, HCN1, CACNA1C, KCTD13, KCNV1, KCNN3, KCNJ13, KCNB1*	1.90×10^{-2}

Note: The table shows GO terms identified by DAVID.[86,87] that are enriched among 302 genes that have a polyevidence score of 2 and above. P_{adj} values represent P-values corrected by the Benjamini-Hochberg procedure in DAVID.

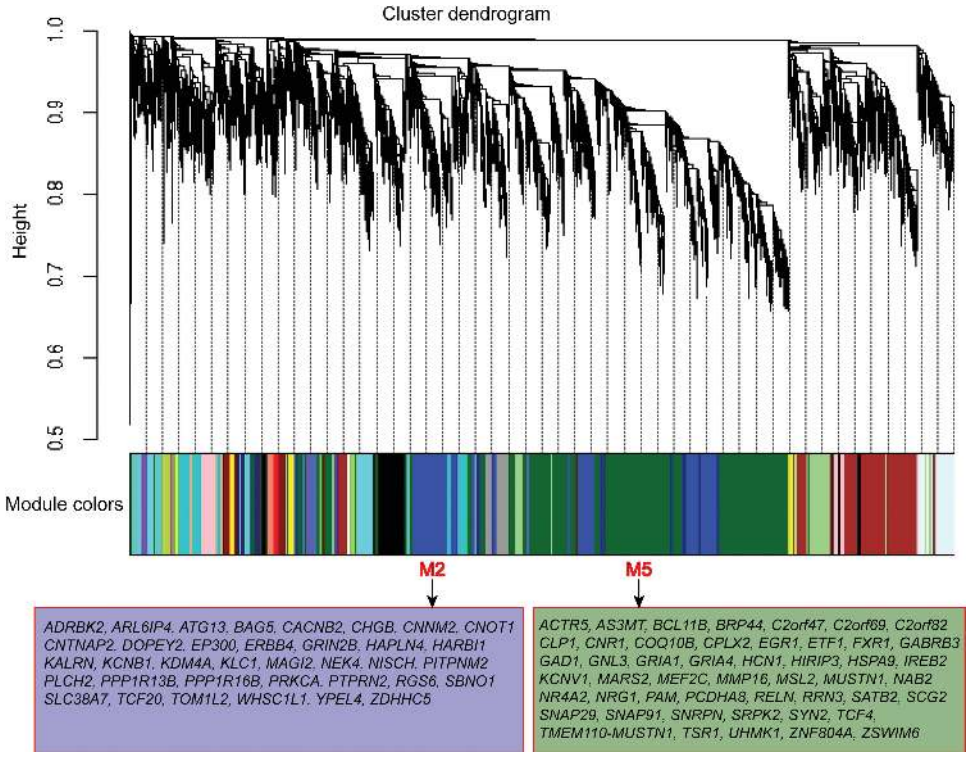

Fig. 3. Co-expression module analysis and enrichment of schizophrenia risk genes in co-expression module 2 and module 5. Schizophrenia candidate genes in module 2 (M2) and module 5 (M5) are shown in box.

interconnected PPI network (figure 4). This result further supported our previous findings[78] and suggested that perturbations of common underlying molecular processes or pathways modulate risk to SZ.

Discussion

In recent years, significant progresses have been made to dissect the genetic architecture of SZ. Multiple highly-significant risk variants have been successfully identified by GWAS.[21–36] Nevertheless, pinpointing the causal risk variant(s) remains a major challenge. In addition to risk genes identified by GWAS, other promising SZ candidate genes have also been revealed by other approaches, such as genetic linkage and association,[18] gene expression study, convergent functional genomics,[42] *Sherlock* integrative analysis.[52] With the increase of sample size and improvement of research approaches, novel SZ risk variants and genes are continuously being identified. Though a large number of risk variants and loci have been identified, the pathophysiology of SZ remains largely unknown. Evidently, it is the time to systematically integrate multi-type data from various studies and to harvest meaningful biological information from genetic findings of SZ. Therefore, we developed the SZDB, a comprehensive database that integrates multi-type resources from different layers of SZ researches, to fill the lacuna.

Compared with other SZ databases such as the SZGene[20] and SZGR,[55] the SZDB has several advantages. First, the SZ risk genes included in the SZGene and SZGR were mainly based on genetic linkage and association studies,

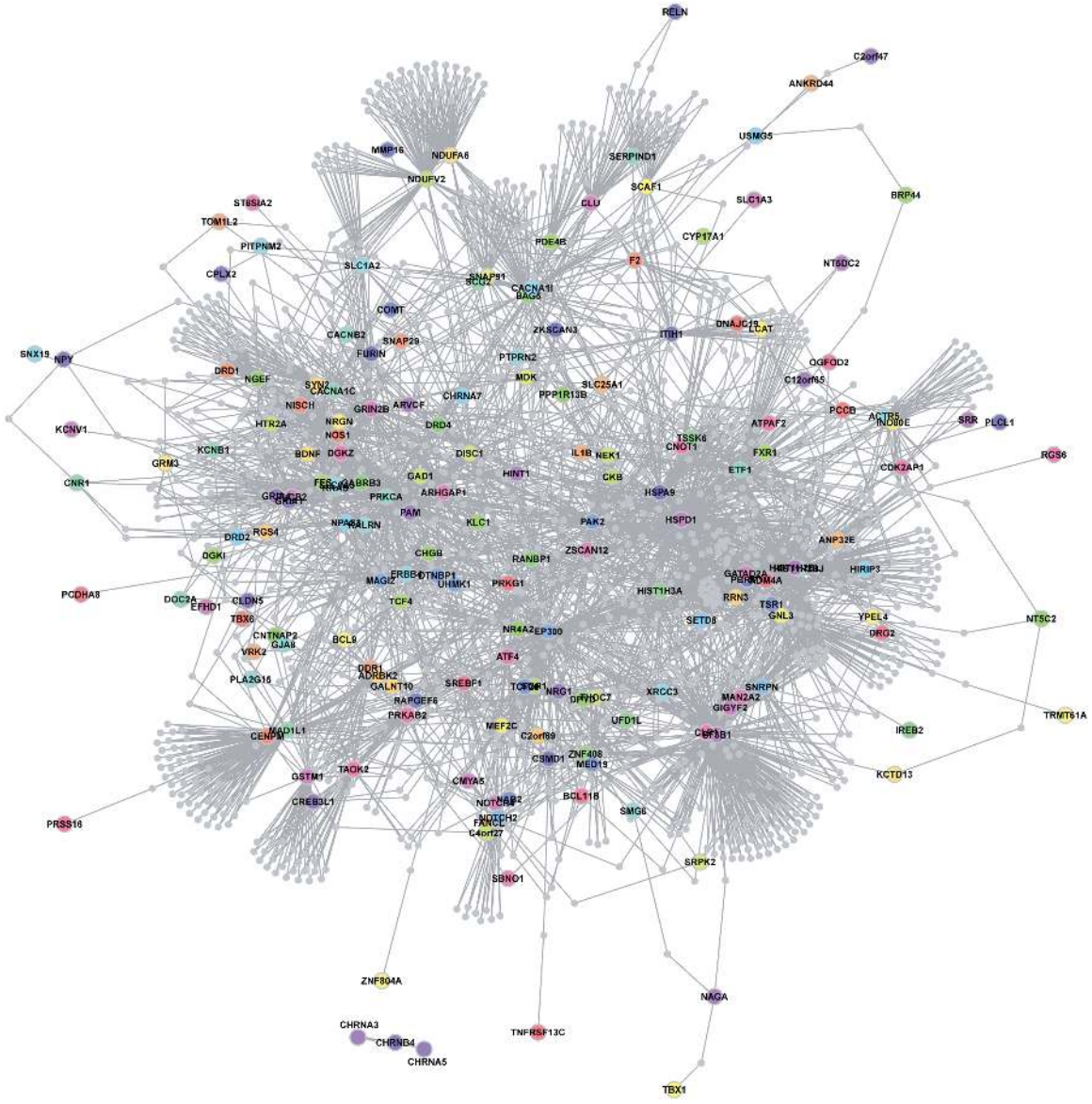

Fig. 4. Proteins encoded by prioritized schizophrenia risk genes form a densely interconnected protein–protein interaction (PPI) network. By using the high-confidence PPI data from InWeb, proteins encoded by prioritized schizophrenia risk genes (that have a final score of 2 and above) were used to construct the PPI network. Proteins encoded by prioritized schizophrenia risk genes are shown.

and these databases were not updated since the establishment in 2008 and 2010, respectively. Moreover, we found that the SZGR could not be even accessible during the reviewing of this manuscript. We noticed that the SZGR was recently moved to a new website (https://bioinfo.uth.edu/SZGR/). We did not check the content of the new website of the SZGR, but genes identified by GWAS and other integrative analyses such as convergent functional genomics[42] and *Sherlock*[52] integrative analysis were not included in the SZGene and the old SZGR databases. Second, the SZDB has a powerful analysis module and users can perform customized analysis, including eQTL, PPI, co-expression, spatio-temporal expression pattern analyses, to name a few. Third, on the basis of multi-type data from various resources, in-depth data integration was conducted and top SZ risk genes were prioritized in the SZDB. Interestingly, we noted that *DRD2* was among the top prioritized genes and had the highest final scores. Dysregulation of the dopamine system function has been well characterized in SZ and almost all available antipsychotic drugs exert their main therapeutic effects though the blockade of DRD2.[88–91] The top prioritized genes thus represent promising SZ risk genes which might be worthy of further functional characterization. Fourth, the SZDB will be updated periodically to keep up with the research progress. With the use of new technologies and analysis methods, novel SZ risk variants and genes are continuously being identified. These latest findings will be integrated into SZDB periodically.

This study has several limitations. First, we used a simple and arbitrary scoring algorithm to prioritize the promising SZ candidate genes from diverse sources (eg, genetic and gene expression studies). Though this scoring system is simple and operational, it ignores the potential overlapping between the sources used for scoring. For example, genes identified by convergent functional genomics (CFG)[42] also integrated information from genetic and gene expression studies. Second, we treated all evidence from different sources equally. Nevertheless, genes identified by GWAS represent the high-confidence SZ candidate genes, therefore, these genes should be given higher weight than genes identified by linkage or association study. These limitations can be addressed through developing an appropriate scoring algorithm in the future.

In conclusion, the SZDB aims to provide a comprehensive resource for SZ genetic research. To the best of our knowledge, SZDB is the most comprehensive SZ genetic database that integrates multiple layers of data from various sources so far. With more genes being identified and more data resources being available, SZDB will be updated every 6 months to maintain the up-to-date resources. The SZDB will integrate more types of data in future, including large-scale data from whole-genome sequencing, tissue-specific PPI interactions, gene expression studies from RNA sequencing. Evidence from epigenetics and animal models will also be integrated into SZDB in future. The ultimate goal of SZDB is to provide a one-stop service for SZ researchers, and they can find most of the information they wanted in the SZDB. With the increase of new genes and integration of more data type, the SZDB will become more powerful. We hope the SZDB will facilitate the translation of genetic findings into molecular risk mechanisms for SZ, and ultimately to the improvement of disease diagnosis and treatment.

Supplementary Material

Supplementary material is available at http://schizophreniabulletin.oxfordjournals.org.

Funding

Strategic Priority Research Program (B) of the Chinese Academy of Sciences (XDB02020003); 1000 Young Talents Program.

Acknowledgments

We thank Dr Long Fan for technical assistance and the 4 reviewers for their critical comments on the early version of the manuscript. The authors declared no potential conflicts of interest.

References

1. Saha S, Chant D, Welham J, McGrath J. A systematic review of the prevalence of schizophrenia. *PLoS Med*. 2005;2:e141.
2. Tandon R, Keshavan MS, Nasrallah HA. Schizophrenia, "Just the Facts": what we know in 2008 part 1: overview. *Schizophr Res*. 2008;100:4–19.
3. World Health Organization. *The Global Burden of Disease: 2004 Update*. Geneva, Switzerland: WHO Press; 2008.
4. Knapp M, Mangalore R, Simon J. The global costs of schizophrenia. *Schizophr Bull*. 2004;30:279–293.
5. Wu EQ, Birnbaum HG, Shi L, et al. The economic burden of schizophrenia in the United States in 2002. *J Clin Psychiatry*. 2005;66:1122–1129.
6. Goeree R, Farahati F, Burke N, et al. The economic burden of schizophrenia in Canada in 2004. *Curr Med Res Opin*. 2005;21:2017–2028.
7. Lieberman JA, Stroup TS, McEvoy JP, et al. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *N Engl J Med*. 2005;353:1209–1223.
8. Lieberman JA. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia: efficacy, safety and cost outcomes of CATIE and other trials. *J Clin Psychiatry*. 2007;68:e04.
9. Walker E, Kestler L, Bollini A, Hochman KM. Schizophrenia: etiology and course. *Annu Rev Psychol*. 2004;55:401–430.
10. Buka SL, Tsuang MT, Lipsitt LP. Pregnancy/delivery complications and psychiatric diagnosis. A prospective study. *Arch Gen Psychiatry*. 1993;50:151–156.
11. Dalman C, Allebeck P, Cullberg J, Grunewald C, Köster M. Obstetric complications and the risk of schizophrenia: a longitudinal study of a national birth cohort. *Arch Gen Psychiatry*. 1999;56:234–240.
12. Barr CE, Mednick SA, Munk-Jorgensen P. Exposure to influenza epidemics during gestation and adult schizophrenia. A 40-year study. *Arch Gen Psychiatry*. 1990;47:869–874.

13. Brown AS, Cohen P, Harkavy-Friedman J, et al. A.E. Bennett Research Award. Prenatal rubella, premorbid abnormalities, and adult schizophrenia. *Biol Psychiatry*. 2001;49:473–486.
14. van Os J, Selten JP. Prenatal exposure to maternal stress and subsequent schizophrenia. The May 1940 invasion of The Netherlands. *Br J Psychiatry*. 1998;172:324–326.
15. Quenstedt M, Parshall A. Prenatal exposure to maternal stress and subsequent schizophrenia. *Br J Psychiatry*. 1998;173:183.
16. Sullivan PF, Kendler KS, Neale MC. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch Gen Psychiatry*. 2003;60:1187–1192.
17. Sullivan PF, Daly MJ, O'Donovan M. Genetic architectures of psychiatric disorders: the emerging picture and its implications. *Nat Rev Genet*. 2012;13:537–551.
18. Lewis CM, Levinson DF, Wise LH, et al. Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: Schizophrenia. *Am J Hum Genet*. 2003;73:34–48.
19. Ng MY, Levinson DF, Faraone SV, et al. Meta-analysis of 32 genome-wide linkage studies of schizophrenia. *Mol Psychiatry*. 2009;14:774–785.
20. Allen NC, Bagade S, McQueen MB, et al. Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the SzGene database. *Nat Genet*. 2008;40:827–834.
21. O'Donovan MC, Craddock N, Norton N, et al. Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nat Genet*. 2008;40:1053–1055.
22. Purcell SM, Wray NR, Stone JL, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009;460:748–752.
23. Ripke S, O'Dushlaine C, Chambert K, et al. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat Genet*. 2013;45:1150–1159.
24. Ikeda M, Aleksic B, Kinoshita Y, et al. Genome-wide association study of schizophrenia in a Japanese population. *Biol Psychiatry*. 2011;69:472–478.
25. Hamshere ML, Walters JT, Smith R, et al. Genome-wide significant associations in schizophrenia to ITIH3/4, CACNA1C and SDCCAG8, and extensive replication of associations reported by the Schizophrenia PGC. *Mol Psychiatry*. 2013;18:708–712.
26. Rietschel M, Mattheisen M, Degenhardt F, et al. Association between genetic variation in a region on chromosome 11 and schizophrenia in large samples from Europe. *Mol Psychiatry*. 2012;17:906–917.
27. Irish Schizophrenia Genomics Consortium & Wellcome Trust Case Control Consortium. Genome-wide association study implicates HLA-C*01:02 as a risk factor at the major histocompatibility complex locus in schizophrenia. *Biol Psychiatry*. 2012;72:620–628.
28. Shi J, Levinson DF, Duan J, et al. Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature*. 2009;460:753–757.
29. Shi Y, Li Z, Xu Q, et al. Common variants on 8p12 and 1q24.2 confer risk of schizophrenia. *Nat Genet*. 2011;43:1224–1227.
30. Stefansson H, Ophoff RA, Steinberg S, et al. Common variants conferring risk of schizophrenia. *Nature*. 2009;460:744–747.
31. Li J, Zhou G, Ji W, et al. Common variants in the BCL9 gene conferring risk of schizophrenia. *Arch Gen Psychiatry*. 2011;68:232–240.
32. Steinberg S, de Jong S, Andreassen OA, et al. Common variants at VRK2 and TCF4 conferring risk of schizophrenia. *Hum Mol Genet*. 2011;20:4076–4081.
33. Wong EH, So HC, Li M, et al. Common variants on Xq28 conferring risk of schizophrenia in Han Chinese. *Schizophr Bull*. 2014;40:777–786.
34. Yue WH, Wang HF, Sun LD, et al. Genome-wide association study identifies a susceptibility locus for schizophrenia in Han Chinese at 11p11.2. *Nat Genet*. 2011;43:1228–1231.
35. Lencz T, Guha S, Liu C, et al. Genome-wide association study implicates NDST3 in schizophrenia and bipolar disorder. *Nat Commun*. 2013;4:2739.
36. Schizophrenia Working Group of the Psychiatric Genomics Consortium*. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 2014;511:421–427.
37. Stefansson H, Rujescu D, Cichon S, et al.; GROUP. Large recurrent microdeletions associated with schizophrenia. *Nature*. 2008;455:232–236.
38. The International Schizophrenia Consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature*. 2008;455:237–241.
39. Xu B, Roos JL, Levy S, et al. Strong association of de novo copy number mutations with sporadic schizophrenia. *Nat Genet*. 2008;40:880–885.
40. Walsh T, McClellan JM, McCarthy SE, et al. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science*. 2008;320:539–543.
41. Malhotra D, McCarthy S, Michaelson JJ, et al. High frequencies of de novo CNVs in bipolar disorder and schizophrenia. *Neuron*. 2011;72:951–963.
42. Ayalew M, Le-Niculescu H, Levey DF, et al. Convergent functional genomics of schizophrenia: from comprehensive understanding to genetic risk prediction. *Mol Psychiatry*. 2012;17:887–905.
43. Horváth S, Mirnics K. Schizophrenia as a disorder of molecular pathways. *Biol Psychiatry*. 2015;77:22–28.
44. Hall J, Trent S, Thomas KL, O'Donovan MC, Owen MJ. Genetic risk for schizophrenia: convergence on synaptic pathways involved in plasticity. *Biol Psychiatry*. 2015;77:52–58.
45. Pruim RJ, Welch RP, Sanna S, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*. 2010;26:2336–2337.
46. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res*. 2012;22:1790–1797.
47. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489:57–74.
48. Xie D, Boyle AP, Wu L, et al. Dynamic trans-acting factor colocalization in human cells. *Cell*. 2013;155:713–724.
49. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res*. 2003;31:3812–3814.
50. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*. 2009;4:1073–1081.
51. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7:248–249.
52. Luo XJ, Mattheisen M, Li M, et al. Systematic Integration of Brain eQTL and GWAS Identifies ZNF323 as a Novel Schizophrenia Risk Gene and Suggests Recent Positive Selection Based on Compensatory Advantage on Pulmonary Function. *Schizophr Bull*. 2015;41:1294–1308.
53. Purcell SM, Moran JL, Fromer M, et al. A polygenic burden of rare disruptive mutations in schizophrenia. *Nature*. 2014;506:185–190.

54. Fromer M, Pocklington AJ, Kavanagh DH, et al. De novo mutations in schizophrenia implicate synaptic networks. *Nature*. 2014;506:179–184.
55. Jia P, Sun J, Guo AY, Zhao Z. SZGR: a comprehensive schizophrenia gene resource. *Mol Psychiatry*. 2010;15:453–462.
56. Sun J, Kuo PH, Riley BP, Kendler KS, Zhao Z. Candidate genes for schizophrenia: a survey of association studies and gene ranking. *Am J Med Genet B Neuropsychiatr Genet*. 2008;147B:1173–1181.
57. Luo X, Huang L, Han L, et al. Systematic prioritization and integrative analysis of copy number variations in schizophrenia reveal key schizophrenia susceptibility genes. *Schizophr Bull*. 2014;40:1285–1299.
58. He X, Fuller CK, Song Y, et al. Sherlock: detecting gene-disease associations by matching patterns of expression QTL and GWAS. *Am J Hum Genet*. 2013;92:667–680.
59. Lanz TA, Joshi JJ, Reinhart V, et al. STEP levels are unchanged in pre-frontal cortex and associative striatum in post-mortem human brain samples from subjects with schizophrenia, bipolar disorder and major depressive disorder. *PLoS One*. 2015;10:e0121744.
60. Iwamoto K, Bundo M, Kato T. Altered expression of mitochondria-related genes in postmortem brains of patients with bipolar disorder or schizophrenia, as revealed by large-scale DNA microarray analysis. *Hum Mol Genet*. 2005;14:241–253.
61. Narayan S, Tang B, Head SR, et al. Molecular profiles of schizophrenia in the CNS at different stages of illness. *Brain Res*. 2008;1239:235–248.
62. Chen C, Cheng L, Grennan K, et al. Two gene co-expression modules differentiate psychotics and controls. *Mol Psychiatry*. 2013;18:1308–1314.
63. de Baumont A, Maschietto M, Lima L, et al. Innate immune response is differentially dysregulated between bipolar disease and schizophrenia. *Schizophr Res*. 2015;161:215–221.
64. Fillman SG, Cloonan N, Catts VS, et al. Increased inflammatory markers identified in the dorsolateral prefrontal cortex of individuals with schizophrenia. *Mol Psychiatry*. 2013;18:206–214.
65. Lamparter D, Marbach D, Rueedi R, Kutalik Z, Bergmann S. Fast and rigorous computation of gene and pathway scores from snp-based summary statistics. *PLoS Comput Biol*. 2016;12:e1004714.
66. Aberg KA, McClay JL, Nerella S, et al. Methylome-wide association study of schizophrenia: identifying blood biomarker signatures of environmental insults. *JAMA Psychiatry*. 2014;71:255–264.
67. Montano C, Taub MA, Jaffe A, et al. Association of DNA methylation differences with schizophrenia in an epigenome-wide association study. *JAMA Psychiatry*. 2016;73:506–514.
68. Hannon E, Spiers H, Viana J, et al. Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci. *Nat Neurosci*. 2016;19:48–54.
69. Myers AJ, Gibbs JR, Webster JA, et al. A survey of genetic human cortical gene expression. *Nat Genet*. 2007;39:1494–1499.
70. 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.
71. BrainSpan. BrainSpan: Atlas of the Developing Human Brain [Internet]. 2013. http://www.brainspan.org. Accessed April 28, 2015.
72. Willsey AJ, Sanders SJ, Li M, et al. Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell*. 2013;155:997–1007.
73. Gulsuner S, Walsh T, Watts AC, et al. Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. *Cell*. 2013;154:518–529.
74. Colantuoni C, Lipska BK, Ye T, et al. Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature*. 2011;478:519–523.
75. Gilman SR, Chang J, Xu B, et al. Diverse types of genetic variation converge on functional gene networks involved in schizophrenia. *Nat Neurosci*. 2012;15:1723–1728.
76. Oti M, Brunner HG. The modular nature of genetic diseases. *Clin Genet*. 2007;71:1–11.
77. Jia P, Zhao Z. Network.assisted analysis to prioritize GWAS results: principles, methods and perspectives. *Hum Genet*. 2014;133:125–138.
78. Luo X, Huang L, Jia P, et al. Protein-protein interaction and pathway analyses of top schizophrenia genes reveal schizophrenia susceptibility genes converge on common molecular networks and enrichment of nucleosome (chromatin) assembly genes in schizophrenia susceptibility loci. *Schizophr Bull*. 2014;40:39–49.
79. Luo XJ, Huang L, Li M, Gan L. Protein-protein interaction analysis reveals common molecular processes/pathways that contribute to risk of schizophrenia. *Schizophr Res*. 2013;143:390–392.
80. Lage K, Karlberg EO, Størling ZM, et al. A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nat Biotechnol*. 2007;25:309–316.
81. Lage K, Hansen NT, Karlberg EO, et al. A large-scale analysis of tissue-specific pathology and gene expression of human disease genes and complexes. *Proc Natl Acad Sci U S A*. 2008;105:20870–20875.
82. Parikshak NN, Luo R, Zhang A, et al. Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell*. 2013;155:1008–1021.
83. Rask-Andersen M, Masuram S, Schiöth HB. The druggable genome: Evaluation of drug targets in clinical trials suggests major shifts in molecular class and indication. *Annu Rev Pharmacol Toxicol*. 2014;54:9–26.
84. Law V, Knox C, Djoumbou Y, et al. DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res*. 2014;42:D1091–D1097.
85. Kang HJ, Kawasawa YI, Cheng F, et al. Spatio-temporal transcriptome of the human brain. *Nature*. 2011;478:483–489.
86. Huang DW, Sherman BT, Tan Q, et al. DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res*. 2007;35:W169–W175.
87. Huang DW, Sherman BT, Tan Q, et al. The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol*. 2007;8:R183.
88. Masri B, Salahpour A, Didriksen M, et al. Antagonism of dopamine D2 receptor/beta-arrestin 2 interaction is a common property of clinically effective antipsychotics. *Proc Natl Acad Sci U S A*. 2008;105:13656–13661.
89. Frégeau MO, Carrier M, Guillemette G. Mechanism of dopamine D2 receptor-induced Ca(2+) release in PC-12 cells. *Cell Signal*. 2013;25:2871–2877.
90. Miyamoto S, Miyake N, Jarskog LF, Fleischhacker WW, Lieberman JA. Pharmacological treatment of schizophrenia: a critical review of the pharmacology and clinical effects of current and future therapeutic agents. *Mol Psychiatry*. 2012;17:1206–1227.
91. Seeman P. Targeting the dopamine D2 receptor in schizophrenia. *Expert Opin Ther Targets*. 2006;10:515–531.