Integrating genome-wide association study with regulatory SNP annotation information identified candidate genes and pathways for schizophrenia

Xiao Liang1,*, Sen Wang1,*, Li Liu1, Yanan Du1, Bolun Cheng1, Yan Wen1, Yan Zhao1, Miao Ding1, Shiqiang Cheng1, Mei Ma1, Lu Zhang1, Xin Qi1, Ping Li1, Xiong Guo1, Feng Zhang1

1Key Laboratory of Trace Elements and Endemic Diseases of National Health and Family Planning Commission, School of Public Health, Health Science Center, Xi’an Jiaotong University, Xi’an, China
*Equal contribution

Correspondence to: Feng Zhang; email: fzhxjtu@mail.xjtu.edu.cn
Keywords: schizophrenia, regulatory SNP, genome-wide association studies, gene ontology (GO), pathways
Received: March 3, 2019 Accepted: May 29, 2019 Published: June 7, 2019

ABSTRACT

Background: Schizophrenia is a complex mental disorder. The genetic mechanism of schizophrenia remains elusive now.

Methods: We conducted a large-scale integrative analysis of two genome-wide association studies of schizophrenia with functional annotation datasets of regulatory single-nucleotide polymorphism (rSNP). The significant SNPs identified by the two genome-wide association studies were first annotated to obtain schizophrenia associated rSNPs and their target genes and proteins, respectively. We then compared the integrative analysis results to identify the common rSNPs and their target regulatory genes and proteins, shared by the two genome-wide association studies of schizophrenia. Finally, DAVID tool was used to conduct gene ontology and pathway enrichment analysis of the identified targets genes and proteins.

Results: We detected 53 schizophrenia-associated target genes for rSNP, such as FOS \((P \text{ value} = 2.18\times10^{-20})\), ATXN1 \((P \text{ value} = 5.22\times10^{-21})\) and HLA-DQA1 \((P \text{ value} = 1.98\times10^{-10})\). Pathway enrichment analysis identified 24 pathways for transcription factors binding regions, chromatin interacting regions, long non-coding RNAs, topologically associated domains, circular RNAs and post-translational modifications, such as hsa05034:Alcoholism \((P \text{ value} = 2.57\times10^{-7})\) and hsa04612:Antigen processing and presentation \((P \text{ value} = 6.82\times10^{-8})\).

Conclusion: We detected multiple candidate genes, gene ontology terms and pathways for schizophrenia, supporting the functional importance of rSNPs, and providing novel clues for understanding the genetic architecture of schizophrenia.

INTRODUCTION

Schizophrenia (SCZ) is a chronic and severe mental disorder, characterized by psychosis, apathy and withdrawal, and cognitive impairment. Depression, anxiety, and substance abuse are additional mental health problems of SCZ condition. The symptoms of SCZ typically appear gradually, start between ages 16 and 30, and never resolve in many cases. About 1% of the population is affected by SCZ during their lifetimes, which is associated with substantial morbidity and mortality, as well as personal and societal costs [1, 2]. Moreover, SCZ is ranked among the top 25 leading causes of disability worldwide [3].
SCZ is a multi-factorial disease, and its occurrence depends on environmental and genetic factors. The high heritability of SCZ points to a major role for inherited genetic variants in the etiology of SCZ, with estimated heritability up to 80% [1, 2]. In recent years, considerable progress has been done in the genetic studies of SCZ [4]. Multiple susceptibility genes have been identified for SCZ [2, 5], such as NRG1 [5], DISC1 [5] and DRD2 [2]. Major histocompatibility complex (MHC) [1] on chromosome 6p is the most significant region for SCZ, which contains many markers reaching genome-wide significance [4]. However, the genetic mechanism of SCZ remains largely unknown.

Genome-wide association studies (GWAS) have great power to identify susceptibility genetic loci associated with complex diseases. Over the years, a number of credible candidate genes for SCZ has been identified, largely by GWAS [2]. Because a stringent threshold requires a huge sample size to reliably identify risk genes, the significant loci identified by GWAS are usually limited and functionally independent, providing limited information for the mechanism studies of SCZ. Moreover, a large part of genetic variants detected by GWAS are located in non-coding chromosomal regions [4]. It is usually confusing how these non-coding regions implicated would be involved in the development of SCZ.

Functional SNPs located in protein-coding and non-coding genes are named regulatory single nucleotide polymorphisms (rSNPs), which usually have major impacts on gene functions [6]. They mainly include transcription factor binding regions (TFBRs), chromatin interactive regions (CIRs), topologically associated domains (TADs), long non-coding RNAs (IncRNAs) coding regions, and circular RNA (circRNA) regions. Additionally, some SNPs located in the protein coding regions can alter protein post-translational modifications (PTMs) [7], such as phosphorylation, methylation, acetylation, ubiquitination, and glycosylation. The implication of rSNPs in the development of complex diseases has been well documented in previous studies [8–10]. Integrating GWAS and functional rSNPs annotation information have improved GWAS power and provided novel clues for the genetic studies of complex diseases [11–13], such as periodontal diseases [13] and breast cancer [12]. To the best of our knowledge, limited efforts have been paid to explore the functional relevance of rSNPs with SCZ.

In this study, we conducted a large-scale integrative analysis of two GWAS datasets of SCZ with functional annotation datasets of rSNPs. The significant SNPs identified by the two GWAS were first annotated to obtain SCZ-associated rSNPs and their target gene and proteins, respectively. We then compared the integrative analysis results to identify the common rSNPs and their target gene and proteins, shared by the two GWAS of SCZ. Finally, DAVID tool was used to conduct gene ontology (GO) and pathway enrichment analysis of the identified target genes and proteins shared by the two GWAS of SCZ.

RESULTS

Analysis results of GWAS and rSNP annotation datasets

For TFBRs, CIRs, IncRNAs regions, TADs and circRNAs, we identified 1,499 SCZ associated rSNPs, corresponding to 35 genes, such as FOS (P value = 2.18×10^{-20}), GABBR1 (P value = 2.18×10^{-20}), MDK (P value = 1.89×10^{-10}) and ATXN1 (P value = 5.22×10^{-21}). For PTMs, we detected 43 rSNPs, corresponding to 18 genes, such as HLA-DQA1 (P value = 1.98×10^{-10}), HLA-DRB1 (P value = 1.36×10^{-12}) and ZSCAN31 (P value = 8.78×10^{-10}) (Table 1).

GO enrichment analysis

GO enrichment analysis identified 15 GO terms enriched in the identified target genes of TFBRs, CIRs, IncRNAs, TADs and circRNAs, such as GO:0000786–nucleosome (P value = 1.84×10^{-10}), GO:0046982–protein heterodimerization activity (P value = 5.97×10^{-7}), GO:0000788–nuclear nucleosome (P value = 5.63×10^{-5}) and GO:0006334–nucleosome assembly (P value = 5.70×10^{-5}). For PTMs, we identified 37 SCZ-associated GO terms, such as GO:0002504–antigen processing and presentation of peptide or polysaccharide antigen via MHC class II (P value = 4.79×10^{-7}), GO:0042613–MHC class II protein complex (P value = 1.03×10^{-10}), GO:0042605–peptide antigen binding (P value = 2.26×10^{-9}) and GO:0071556–integral component of lumenal side of endoplasmic reticulum membrane (P value = 2.43×10^{-6}) (Table 2).

Pathway enrichment analysis

For TFBRs, CIRs, IncRNAs, TADs and circRNAs, we identified 3 pathways associated with SCZ, including ha05322:Systemic lupus erythematosus (P value = 3.77×10^{-8}), ha05034:Alcoholism (P value = 2.57×10^{-7}) and ha05203:Viral carcinogenesis (P value = 1.78×10^{-2}). For PTMs, we identified 21 pathways associated with SCZ, such as ha04612:Antigen processing and presentation (P value = 6.82×10^{-9}), ha05310:Asthma (P value = 7.44×10^{-7}), ha05332:Graft-versus-host disease (P value = 1.00×10^{-6}) and ha04672:Intestinal
Table 1. List of SCZ associated rSNPs and their target regulatory genes and proteins shared by both SCZ 1 and SCZ 2.

| SNP          | Gene    | SNP-related regulatory elements                                                                 | P value   |
|--------------|---------|--------------------------------------------------------------------------------------------------|-----------|
| rs35001169   | HIST1H3J | TFBRs, CIRs, IncRNAs regions, TADs, circRNAs                                                      | $7.40 \times 10^{-27}$ |
| rs35001169   | HIST1H2AM | TFBRs, CIRs, IncRNAs regions, TADs, circRNAs                                                      | $7.40 \times 10^{-27}$ |
| rs35819751   | MIR3143  | TFBRs, CIRs, IncRNAs regions, TADs, circRNAs                                                      | $9.52 \times 10^{-27}$ |
| rs66462181   | HIST1H4A | TFBRs, CIRs, IncRNAs regions, TADs, circRNAs                                                      | $1.06 \times 10^{-27}$ |
| rs17695758   | DNAH8    | TFBRs, CIRs, IncRNAs regions, TADs, circRNAs                                                      | $1.52 \times 10^{-26}$ |
| rs141342723  | HIST1H2BL| TFBRs, CIRs, IncRNAs regions, TADs, circRNAs                                                      | $1.82 \times 10^{-26}$ |
| rs13209332   | HIST1H2AK| TFBRs, CIRs, IncRNAs regions, TADs, circRNAs                                                      | $2.10 \times 10^{-26}$ |
| rs2232423    | ZSCAN12  | PTMs                                                                                             | $1.71 \times 10^{-23}$ |
| rs33932084   | PGBD1    | PTMs                                                                                             | $5.00 \times 10^{-23}$ |
| rs41266839   | BTN3A1   | PTMs                                                                                             | $4.77 \times 10^{-22}$ |
| rs34788973   | OR2B2    | PTMs                                                                                             | $1.87 \times 10^{-21}$ |
| rs34197618   | ATXN1    | TFBRs, CIRs, IncRNAs regions, TADs, circRNas                                                      | $5.22 \times 10^{-21}$ |
| rs41266779   | FOS      | TFBRs, CIRs, IncRNAs regions, TADs, circRNas                                                      | $2.18 \times 10^{-20}$ |
| rs41266779   | HIST1H2BK| TFBRs, CIRs, IncRNAs regions, TADs, circRNas                                                      | $2.18 \times 10^{-20}$ |
| rs41266779   | NUP153   | TFBRs, CIRs, IncRNAs regions, TADs, circRNas                                                      | $2.18 \times 10^{-20}$ |
| rs41266779   | PKHD1    | TFBRs, CIRs, IncRNAs regions, TADs, circRNas                                                      | $2.18 \times 10^{-20}$ |
| rs41266779   | CLIC5    | TFBRs, CIRs, IncRNAs regions, TADs, circRNas                                                      | $2.18 \times 10^{-20}$ |
| rs41266779   | DCDC2    | TFBRs, CIRs, IncRNAs regions, TADs, circRNas                                                      | $2.18 \times 10^{-20}$ |
| rs41266779   | GABBR1   | TFBRs, CIRs, IncRNAs regions, TADs, circRNas                                                      | $2.18 \times 10^{-20}$ |
| rs41266779   | PRIM2    | TFBRs, CIRs, IncRNAs regions, TADs, circRNas                                                      | $2.18 \times 10^{-20}$ |
| rs41266779   | ANKH     | TFBRs, CIRs, IncRNAs regions, TADs, circRNas                                                      | $2.18 \times 10^{-20}$ |
| rs41266779   | HIST1H2AH| TFBRs, CIRs, IncRNAs regions, TADs, circRNas                                                      | $2.18 \times 10^{-20}$ |
| rs35050608   | MBOAT1   | TFBRs, CIRs, IncRNAs regions, TADs, circRNas                                                      | $2.37 \times 10^{-20}$ |
| rs35506517   | LY86     | TFBRs, CIRs, IncRNAs regions, TADs, circRNas                                                      | $3.25 \times 10^{-20}$ |
| rs9393718    | HIST1H2BJ| TFBRs, CIRs, IncRNAs regions, TADs, circRNas                                                      | $3.32 \times 10^{-20}$ |
| rs35555795   | BTN1A1   | PTMs                                                                                             | $9.83 \times 10^{-17}$ |
| rs79780963   | RNU1-60P | TFBRs, CIRs, IncRNAs regions, TADs, circRNas                                                      | $7.51 \times 10^{-16}$ |
| rs13216828   | BTN3A2   | PTMs                                                                                             | $1.59 \times 10^{-15}$ |
| rs9986596    | ZKSCAN4  | PTMs                                                                                             | $5.51 \times 10^{-15}$ |
| rs3891176    | HLA-DQB1 | PTMs                                                                                             | $3.29 \times 10^{-14}$ |
| rs11693528   | BMPR2    | TFBRs, CIRs, IncRNAs regions, TADs, circRNas                                                      | $5.62 \times 10^{-14}$ |
| rs16891235   | HIST1H1A | PTMs                                                                                             | $1.59 \times 10^{-13}$ |
| rs769949     | PLCL1    | TFBRs, CIRs, IncRNAs regions, TADs, circRNas                                                      | $2.10 \times 10^{-13}$ |
| rs853678     | ZSCAN31  | PTMs                                                                                             | $3.42 \times 10^{-13}$ |
| rs281886     | MPP4     | TFBRs, CIRs, IncRNAs regions, TADs, circRNas                                                      | $3.92 \times 10^{-13}$ |
| rs281876     | AOX1     | TFBRs, CIRs, IncRNAs regions, TADs, circRNas                                                      | $3.92 \times 10^{-13}$ |
| rs35220450   | RAPH1    | TFBRs, CIRs, IncRNAs regions, TADs, circRNas                                                      | $4.05 \times 10^{-13}$ |
| rs35220450   | INO80D   | TFBRs, CIRs, IncRNAs regions, TADs, circRNas                                                      | $4.05 \times 10^{-13}$ |
| rs281760     | AOX2P    | TFBRs, CIRs, IncRNAs regions, TADs, circRNas                                                      | $4.71 \times 10^{-13}$ |
| rs10734901   | ATP6V0A2 | TFBRs, CIRs, IncRNAs regions, TADs, circRNas                                                      | $4.83 \times 10^{-13}$ |
| rs3098341    | BOLL     | TFBRs, CIRs, IncRNAs regions, TADs, circRNas                                                      | $4.89 \times 10^{-13}$ |
| rs10431750   | KLC1     | TFBRs, CIRs, IncRNAs regions, TADs, circRNas                                                      | $5.69 \times 10^{-13}$ |
| Term               | Term Description                                                                 | SNP-related regulatory elements | P value      |
|--------------------|----------------------------------------------------------------------------------|---------------------------------|--------------|
| GO:0000786        | nucleosome                                                                        | TFBRs, CIRs, lncRNAs regions, TADs, circRNAs | 1.84×10⁻¹⁰  |
| GO:0002504        | Antigen processing and presentation of peptide or polysaccharide antigen          | TFBRs, CIRs, lncRNAs regions, TADs, circRNAs | 4.79×10⁻⁷    |
| GO:0046982        | protein heterodimerization activity                                               | TFBRs, CIRs, lncRNAs regions, TADs, circRNAs | 5.97×10⁻⁷    |
| GO:0042613        | MHC class II protein complex                                                      | TFBRs, CIRs, lncRNAs regions, TADs, circRNAs | 1.03×10⁻⁶    |
| GO:0042605        | Peptide antigen binding                                                           | TFBRs, CIRs, lncRNAs regions, TADs, circRNAs | 2.26×10⁻⁶    |
| GO:0071556        | Integral component of lumenal side of endoplasmic reticulum membrane              | TFBRs, CIRs, lncRNAs regions, TADs, circRNAs | 2.43×10⁻⁶    |
| GO:0030658        | Transport vesicle membrane                                                        | TFBRs, CIRs, lncRNAs regions, TADs, circRNAs | 5.57×10⁻⁶    |
| GO:0030669        | Clathrin coated endocytic vesicle membrane                                        | TFBRs, CIRs, lncRNAs regions, TADs, circRNAs | 7.03×10⁻⁶    |
| GO:0002381        | Immunoglobulin production involved in immunoglobulin mediated immune response    | PTMs                            | 8.50×10⁻⁶    |
| GO:0050852        | T cell receptor signaling pathway                                                 | PTMs                            | 9.71×10⁻⁶    |
| GO:0012507        | ER to Golgi transport vesicle membrane                                            | PTMs                            | 1.45×10⁻⁵    |
| GO:0002455        | Humoral immune response mediated by circulating immunoglobulin                    | PTMs                            | 1.78×10⁻⁵    |
| GO:0019882        | Antigen processing and presentation                                              | PTMs                            | 1.81×10⁻⁵    |
| GO:0030666        | Endocytic vesicle membrane                                                        | PTMs                            | 2.98×10⁻⁵    |
| GO:0060333        | Interferon gamma mediated signaling pathway                                       | PTMs                            | 3.90×10⁻⁵    |
| GO:0031295        | T cell costimulation                                                              | PTMs                            | 5.17×10⁻⁵    |
| GO:0000788        | nuclear nucleosome                                                                | TFBRs, CIRs, lncRNAs regions, TADs, circRNAs | 5.63×10⁻⁵    |
| GO:0006334        | nucleosome assembly                                                                | TFBRs, CIRs, lncRNAs regions, TADs, circRNAs | 5.70×10⁻⁵    |
| GO:0032588        | Trans Golgi network membrane                                                      | PTMs                            | 5.92×10⁻⁵    |
| GO:0019883        | Antigen processing and presentation of exogenous peptide antigen via MHC class II | PTMs                            | 8.46×10⁻⁵    |
| GO:0032395        | MHC class II receptor activity                                                    | PTMs                            | 8.78×10⁻⁵    |
| GO:0003677        | DNA binding                                                                       | TFBRs, CIRs, lncRNAs regions, TADs, circRNAs | 2.20×10⁻⁴    |
| GO:0005737        | cytoplasm                                                                         | TFBRs, CIRs, lncRNAs regions, TADs, circRNAs | 5.83×10⁻⁴    |

Note: TFBRs, transcription factor binding regions; CIRs, chromatin interactive regions; TADs, topologically associated domains; PTMs, protein post-translational modifications.

Table 2. List of SCZ associated gene ontology terms shared by both SCZ 1 and SCZ 2.
| GO:0043565 | Sequence specific DNA binding | PTMs | 1.19×10^{-3} |
|-------------|-------------------------------|------|--------------|
| GO:0005765  | Lysosomal membrane            | PTMs | 1.96×10^{-3} |
| GO:2001179  | Regulation of interleukin 10 secretion | PTMs | 2.86×10^{-3} |
| GO:0006342  | chromatin silencing           |     |              |
| GO:0032673  | Regulation of interleukin 4 production | PTMs | 3.81×10^{-3} |
| GO:0072643  | Interferon gamma secretion    | PTMs | 6.65×10^{-3} |
| GO:0006955  | Immune response               | PTMs | 6.87×10^{-3} |
| GO:003700   | Transcription factor activity, sequence specific DNA binding | PTMs | 1.10×10^{-2} |
| GO:0042088  | T helper 1 type immune response Membrane | PTMs | 1.14×10^{-2} |
| GO:0016020  | Detection of bacterium        | PTMs | 1.17×10^{-2} |
| GO:0016045  | Inflammatory response to antigenic stimulus | PTMs | 1.23×10^{-2} |
| GO:000139   | Golgi membrane                | PTMs | 1.64×10^{-2} |
| GO:0031047  | Gene silencing by RNA         |     |              |
| GO:0042405  | nuclear inclusion body        |     |              |
| GO:0032689  | Negative regulation of interferon gamma production | PTMs | 2.64×10^{-2} |
| GO:0035774  | Positive regulation of insulin secretion involved in cellular response to glucose stimulus | PTMs | 2.73×10^{-2} |
| GO:0016021  | Integral component of membrane | PTMs | 2.81×10^{-2} |
| GO:0010507  | Negative regulation of autophagy | PTMs | 3.38×10^{-2} |
| GO:0007040  | Lysosome organization         | PTMs | 3.38×10^{-2} |
| GO:0042130  | Negative regulation of T cell proliferation | PTMs | 3.47×10^{-2} |
| GO:0051262  | Protein tetramerization       | PTMs | 3.75×10^{-2} |
| GO:0005654  | nucleoplasm                   |     |              |
| GO:0005887  | Integral component of plasma membrane |     |              |
| GO:0007214  | gamma-aminobutyric acid signaling pathway |     |              |
| GO:0070062  | extracellular exosome         |     |              |
| GO:0000790  | nuclear chromatin             |     |              |
| GO:0002227  | innate immune response in mucosa telomere organization |     |              |
| GO:0032200  |                             |     |              |

Note: TFBRs, transcription factor binding regions; CIRs, chromatin interactive regions; TADs, topologically associated domains; PTMs, protein post-translational modifications.

immune network for IgA production \((P \text{ value } = 2.96 \times 10^{-6})\) (Table 3).

**DISCUSSION**

Considering that most of SCZ variants identified by GWAS were not causal, integration of the GWAS results with functional rSNPs information is a powerful approach to discover novel candidate genes for SCZ [4]. To evaluate the roles of rSNPs in the pathogenesis of SCZ, we conducted a large-scale integrative genomics analysis of two GWAS datasets of SCZ with functional annotation datasets of rSNPs. We identified multiple candidate genes, GO terms, and biological pathways for SCZ. Our study results support the functional importance of rSNPs in the genetic mechanism of SCZ, and provide novel clues for understanding the genetic architecture of SCZ.
We identified several candidate genes for SCZ, such as FOS, GABBR1, MDK, ATXN1 and ZSCAN31. FOS is classified as one of the immediate early genes (IEGs), which encode not only transcription factors, but also a much wider variety of proteins, including signaling molecules, growth factors and cytoskeletal proteins [14]. Alteration in the expression of IEGs is linked to neuron generation and neuronal cell death. Nadia Cattane et al. have reported that FOS was significantly upregulated in the fibroblasts of SCZ patients [14]. Defects in synaptic plasticity are involved in the pathophysiology of SCZ. Interestingly, SNPs either protective (rs1063169) or positively associated with SCZ (rs7101T) were identified, showing transcription factor c-fos was important in the regulation of synaptic plasticity [15]. Huang et al. observed high FOS expression in non-neural peripheral samples and low FOS expression in the brain tissues of SCZ patients compared with healthy controls, which suggests that FOS is a sensitive marker for SCZ [16]. In addition, detection of FOS in blood samples may be helpful for SCZ diagnosis [16]. These combined results support the functional relevance of FOS with SCZ [14–16], which is consistent with our study results.

GABBR1 is another SCZ-associated gene identified by this study. \( \gamma \)-aminobutyric acid B (GABA\(B\)) is an inhibitory transmitter molecule acting at neuronal synapses. Functional GABA\(B\) receptor requires both the GABBR1 and GABBR2 subunits. GABA\(B\) receptor can modulate the release of a number of neurotransmitters, including dopamine, serotonin, noradrenaline, somatostatin, glutamate and GABA [17]. Interestingly, the reduction of GABBR1 in pyramidal cells, and consequent reduction of GABAB receptors, can result in the dysfunction of inhibitory mechanism and increase signal output [18]. Previous studies have also observed association between GABBR1 and SCZ. Fatemi et al. observed a significant reduction of GABBR1 protein in the lateral cerebellum of the subjects with SCZ, bipolar disorder, and major depression [17]. In addition, Zai et al. conducted a case-control study and detected a positive association between GABBR1 and SCZ [19].

Genetic variation in a region on chromosome 11 that contains MDK was significantly associated with SCZ [20]. In addition, MDK also accumulated in senile plaques in the hippocampus of patients with Alzheimer’s disease [20, 21]. Notably, ATXN1 serves as one of the hub genes in the protein-protein interaction network for immunoglobulin A (IgA) production.
The involvement of cell adhesion molecules (CAMs) in the pathophysiology of SCZ has long been hypothesized. In this study, CAMs (hsa04514) pathway was detected for SCZ. The CAMs pathway is implicated in a variety of neurocognitive processes, including memory and attention-related reaction time. Multiple CAM genes has been reported to be associated with SCZ risk [26]. Neural CAMs are very important members of the exclusive group of the molecules responsible for precise wiring of nervous system. Neural CAMs serve as “glue” in cell-to-cell adhesion and contact-mediated attraction [33]. They can interact with numerous matrix components, and are involved in many aspects of neuronal development, synaptogenesis, and myelination, which guide brain morphology and support highly-coordinated brain activity [33].

Another interesting SCZ associated pathway is alcoholism (hsa05034). Recent studies have suggested that alcoholism has a wide-reaching influence on the clinical course of SCZ, contributing to shape abnormalities in hippocampus and subcortical shape differences [38]. We also observed that systemic lupus erythematosus (SLE) (hsa05322) was associated with SCZ. Despite the fact that SCZ is not classified as a typical autoimmune diseases, the dysregulation of the immune system cannot be excluded [39, 40]. Interestingly, DNA and myelin basic protein (MBP)-hydrolyzing antibodies, which play an important harmful role in SLE pathogenesis, were also detected in the sera of SCZ patients. In addition, light chains of IgGs from SCZ patients were similar to those of SLE patients [41].

The majority of SNPs identified by GWAS are enriched in non-coding regions [4], and contribute to complex traits and diseases through various molecular mechanisms. These include effects on transcription factor binding affinities, which can result in differential gene expression [11]. However, significant loci identified by GWAS have rarely been tracked to causal polymorphisms thus far. Integrative analysis of GWAS with functional rSNPs is helpful to improve GWAS

interaction network containing many known SCZ risk genes [22]. Actually, ATXN1 is highly expressed in prefrontal cortex, and SCZ patients had significantly decreased ATXN1 expression [22]. In addition, Takeo Saito et al. have revealed an association of rs7779855 in ZSCAN31 with SCZ [23].

GO analysis detected several GO terms for SCZ. One important finding of this study is the disclosure of the MHC class II protein complex (GO:00042613), a class of MHC molecules like human leukocyte antigens HLA-DQα1 and HLA-DQB1, which present antigens to CD4-positive T-lymphocytes. The association between SCZ and the immune system has been repeatedly revealed in genetic, epidemiological and post mortem studies [24]. The protein encoded by HLA-DQα1 gene binds to the protein encoded by HLA-DQB1. Together, they form a functional protein complex called an antigen-binding DQαβ heterodimer, which displays foreign peptides to immune system to trigger body's immune response. Interestingly, HLA alleles were previously shown to be associated with SCZ risk [25]. rs9272105 within HLA-DQA1 explained 1–3% of variation in attentional control, and to a lesser extent, premorbid intelligence quotient in psychotic patients [26]. Additionally, rs9272105 within HLA-DQA1 was also individually associated with variation in neuropsychological function [26]. It has been demonstrated that the level of HLA-DRB1 in SCZ was decreased in peripheral blood samples in contrast with increased level in prefrontal cortex [27]. Moreover, HLA-DPA1 and CD74, which are integral components of the MHC Class II protein complex, were both reduced in hippocampus, amygdala, and dorsolateral prefrontal cortex regions in SCZ [28].

T helper 1 type immune response (GO:0042088) is another significant GO term detected by this study. Delayed hypersensitivity reaction is the classic cell-mediated immune response with sensitized T helper-1 cells. Michael et al. found an attenuated skin reactivity to various antigens in SCZ patients [29]. They also observed significantly diminished responses of schizophrenics to antigen stimulation with tetanus and diptheria antigen presentation [29]. Significantly increased serum level of interleukin-18, a cytokine known to activate T helper 1 type cells in immune responses, has been previously observed in SCZ patients [30].

In addition, we also identified several other GO terms associated with SCZ, such as Golgi membrane (GO:0000139) and GABA signaling pathway (GO:0007214). Previously, it was shown that differentially expressed genes related to Golgi apparatus, vesicular transport and membrane association were over-represented in SCZ [31]. In line, Devor et al. identified a large number of GABA-associated genes for SCZ [32].
power and provide novel clues for pathogenetic studies of SCZ. Notably, the functional SNPs data can assist to exclude unlikely genes/loci, effectively reducing the number of tests needed for unbiased searches across the genome, thus, improving the power to discover novel causal loci [4]. To the best of our knowledge, this is the first large-scale integrative analysis of GWAS and rSNPs for SCZ. The implication of rSNPs in the development of SCZ was systematically investigated considering TFBRs, CIRs, lncRNAs regions, TADs, circRNAs, and PTMs in our study. Nevertheless, one limitation of this study should be noted. The SCZ-associated SNP sets were driven from previous GWAS. The accuracy of our integrative analysis may be affected by the power of previous GWAS of SCZ. Therefore, further studies are warranted to confirm our findings.

In conclusion, we conducted a large-scale integrative genomics analysis of two GWAS datasets of SCZ with functional annotation datasets of rSNPs to explore the genetic basis of rSNPs in the pathogenesis of SCZ. We observed multiple candidate genes, GO terms and pathways for SCZ. We hope that our study results could provide novel clues for the pathogenic and therapeic studies of SCZ.

MATERIALS AND METHODS

The first GWAS dataset of SCZ (SCZ1)

A large GWAS meta-analysis data of SCZ was driven from the Psychiatric Genomics Consortium (PGC) [42], totally containing 33,426 SCZ cases and 32,541 controls. Genotypes from all studies were processed by the PGC using unified quality control procedures followed by imputation of SNPs and insertion-deletions using the 1000 Genomes Project reference panel [43]. Quality control and imputation were performed on each of the study cohort datasets, according to the standard procedures established by the PGC [42]. Genotype imputation was performed using the pre-phasing/imputation stepwise approach implemented in IMPUTE2 [44] and SHAPEIT [45]. The imputation reference set consists of 2,186 phased haplotypes from the full 1000 Genomes Project dataset. Logistic regression was conducted to control for 13 components of ancestry, study sites and genotyping platform. Detailed description of sample characteristics, experimental design, statistical analysis and quality control can be found in the previous studies [42].

The second GWAS dataset of SCZ (SCZ2)

Another independent GWAS data of SCZ [2] was used here. Briefly, this GWAS included 36,989 SCZ cases and 113,075 controls, from 49 ancestry matched, non-overlapping case-control samples (46 of European and three of East Asian ancestry, 34,241 cases and 45,604 controls) and 3 family-based samples of European ancestry (1,235 parent affected-offspring trios). Genotypes from all studies were processed by the PGC using unified quality control procedures. The 1000 Genomes Project reference panel was used for SNPs imputation [43]. In each sample, association testing was conducted using imputed SNP dosages and principal components to control for population stratification. After quality control (imputation INFO score ≥ 0.6, MAF ≥ 0.01, and successfully imputed in ≥ 20 samples), they considered around 9.5 million variants. An inverse-weighted fixed effects model was used for final meta-analysis. Detailed description of sample characteristics, experimental design, statistical analysis and quality control can be found in the previous study [2].

rSNPs annotation datasets

The rSNPs annotation information were driven from the rSNPBase 3.1 database (http://rsnp3.psych.ac.cn) [46] and the AWESOME database (http://www.awesome-hust.com) [7]. rSNPBase 3.1 provided rich functional annotation for human SNP-related regulatory elements and their target regulatory genes, including TFBRs, CIRs, mature microRNA (miRNA) regions, predicted miRNA target sites, lncRNA regions, TADs and circRNAs. AWESOME database is an analysis tool that systematically evaluates the role of SNPs on nearly all kinds of PTMs based on 20 available tools. They construct a comprehensive platform to collect and integrate SNPs and multiple PTM information, utilizing 24 published database or tools. 1,043,608 germline missense variants from the dbSNP was used and each SNP was matched with its protein sequence in AWESOME. Detailed description of the two rSNPs annotation database can be found in the published studies [7, 46].

Statistical analysis

The significant SNPs with GWAS P value < 5 × 10^{-8} were selected from the two GWAS of SCZ (SCZ1 and SCZ2). The selected SCZ-associated SNPs were then annotated by the rSNPBase 3.1 database [46] and the AWESOME database to obtain SCZ associated rSNPs and their target regulatory genes and proteins. We then compared the integrative analysis results to identify the common rSNPs and their target genes and proteins shared by the two GWAS of SCZ. To explore the functional relevance of identified target regulatory genes and proteins with SCZ, GO and pathway enrichment analyses of the identified common target genes and proteins shared by the SCZ1 and SCZ2 were performed.
by the Database for Annotation, Visualization and Integrated Discovery (DAVID) tool [47].

Abbreviations

GWAS: genome wide association studies; rSNP: regulatory SNP; TFBRs: transcription factor binding regions; CIRs: chromatin interactive regions; IncRNAs: long non-coding RNA regions; TADs: topologically associated domains; circRNAs: circular RNAs; PTMs: protein post-translational modifications; GO: gene ontology; SCZ: Schizophrenia.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

FUNDING

This work was supported by the National Natural Scientific Foundation of China (81673112); the Key projects of international cooperation among governments in scientific and technological innovation (2016YFE0119100); the Natural Science Basic Research Plan in Shaanxi Province of China (2017JZ024); and the Fundamental Research Funds for the Central Universities.

REFERENCES

1. Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, Sklar P, and International Schizophrenia Consortium. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature. 2009; 460:748–52. https://doi.org/10.1038/nature08185 PMID:19571811

2. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. Nature. 2014; 511:421–27. https://doi.org/10.1038/nature13595 PMID:25056061

3. Chong HY, Teoh SL, Wu DB, Kotirum S, Chiou CF, Chaiyakunapruk N. Global economic burden of schizophrenia: a systematic review. Neuropsychiatr Dis Treat. 2016; 12:357–73. https://doi.org/10.2147/ndt.s96649 PMID:26937191

4. Chen J, Cao F, Liu L, Wang L, Chen X. Genetic studies of schizophrenia: an update. Neurosci Bull. 2015; 31:87–98. https://doi.org/10.1007/s12264-014-1494-4 PMID:25652814

5. Craddock N, O'Donovan MC, Owen MJ. Genes for schizophrenia and bipolar disorder? Implications for psychiatric nosology. Schizophr Bull. 2006; 32:9–16. https://doi.org/10.1093/schbul/sbj033 PMID:16319375

6. Ramírez-Bello J, Jiménez-Morales M. [Functional implications of single nucleotide polymorphisms (SNPs) in protein-coding and non-coding RNA genes in multifactorial diseases]. Gac Med Mex. 2017; 153:238–50. https://doi.org/10.1186/1471-2156-9-10 PMID:28474710

7. Yang Y, Peng X, Ying P, Tian J, Li J, Ke J, Zhu Y, Gong Y, Zou D, Yang N, Wang X, Mei S, Zhong R, et al. AWESOME: a database of SNPs that affect protein post-translational modifications. Nucleic Acids Res. 2019; 47:D874–D880. https://doi.org/10.1093/nar/gky821 PMID:30215764

8. Prokunina L, Alarcón-Riquelme ME. Regulatory SNPs in complex diseases: their identification and functional validation. Expert Rev Mol Med. 2004; 6:1–15. https://doi.org/10.1017/S1462399404007690 PMID:15122975

9. Butler MG, McGuire AB, Masoud H, Manzardo AM. Currently recognized genes for schizophrenia: high-resolution chromosome ideogram representation. Am J Med Genet B Neuropsychiatr Genet. 2016; 171B:181–202. https://doi.org/10.1002/ajmg.b.32391 PMID:26462458

10. Wong KC. DNA Motif Recognition Modeling from Protein Sequences. iScience. 2018; 7:198–211. https://doi.org/10.1016/j.isci.2018.09.003 PMID:30267681

11. Cavalli M, Pan G, Nord H, Wallén Arzt E, Wallerman O, Wadelius C. Allele-specific transcription factor binding in liver and cervix cells unveils many likely drivers of GWAS signals. Genomics. 2016; 107:248–54. https://doi.org/10.1016/j.ygeno.2016.04.006 PMID:27126307

12. Liu Y, Walavalkar NM. Identification of breast cancer associated variants that modulate transcription factor binding. PLoS Genet. 2017; 13:e1006761. https://doi.org/10.1371/journal.pgen.1006761 PMID:28957321

13. Razzouk S. Regulatory elements and genetic variations in periodontal diseases. Arch Oral Biol. 2016; 72:106–15. https://doi.org/10.1016/j.archoralbio.2016.08.015 PMID:27569042
14. Cattane N, Minelli A, Milanesi E, Maj C, Bignotti S, Bortolomasi M, Bocchio Chiavetto L, Gennarelli M. Altered gene expression in schizophrenia: findings from transcriptional signatures in fibroblasts and blood. PLoS One. 2015; 10:e0116686. https://doi.org/10.1371/journal.pone.0116686 PMID:25658856

15. Boyajyan A, Zakharyan R, Atshemyan S, Chavushyan A, Mkrtchyan G. Schizophrenia-associated Risk and Protective Variants of c-Fos Encoding Gene. Recent Adv DNA Gene Seq. 2015; 9:51–57. https://doi.org/10.2174/23520922096661502231133 PMID:25706621

16. Huang J, Liu F, Wang B, Tang H, Teng Z, Li L, Qiu Y, Wu H, Chen J. Central and Peripheral Changes in FOS Expression in Schizophrenia Based on Genome-Wide Gene Expression. Front Genet. 2019; 10:232. https://doi.org/10.3389/fgene.2019.00232 PMID:30967896

17. Fatemi SH, Folsom TD, Thuras PD. Deficits in GABA(B) receptor system in schizophrenia and mood disorders: a postmortem study. Schizophr Res. 2011; 128:37–43. https://doi.org/10.1016/j.schres.2010.12.025 PMID:21303731

18. Mizukami K, Ishikawa M, Hidaka S, Iwakiri M, Sasaki M, Iritani S. Immunohistochemical localization of GABAB receptor in the entorhinal cortex and inferior temporal cortex of schizophrenic brain. Prog Neuropsychopharmacol Biol Psychiatry. 2002; 26:393–96. https://doi.org/10.1016/S0278-5846(01)00247-0 PMID:11817519

19. Zai G, King N, Wong GW, Barr CL, Kennedy JL. Possible association between the gamma-aminobutyric acid type B receptor 1 (GABBR1) gene and schizophrenia. Eur Neuropsychopharmacol. 2005; 15:347–52. https://doi.org/10.1016/j.euroneuro.2004.12.006 PMID:15820424

20. Rietschel M, Mattheisen M, Degenhardt F, Mühleisen TW, Kirsch P, Esslinger C, Herm S, Demontis D, Steffens M, Strohmaier J, Haenisch B, Breuer R, Czerski PM, et al, and Genetic Risk and Outcome in Psychosis (GROUP Investigators), and SGENE-plus Consortium. Association between genetic variation in a region on chromosome 11 and schizophrenia in large samples from Europe. Mol Psychiatry. 2012; 17:906–17. https://doi.org/10.1038/mp.2011.80 PMID:21747397

21. Yasuhara O, Muramatsu H, Kim SU, Muramatsu T, Maruta H, McGeer PL. Midkine, a novel neurotrophic factor, is present in senile plaques of Alzheimer disease. Biochem Biophys Res Commun. 1993; 192:246–51. https://doi.org/10.1006/bbrc.1993.1406 PMID:8476427

22. Liu J, Su B. Integrated analysis supports ATXN1 as a schizophrenia risk gene. Schizophr Res. 2018; 195:298–305. https://doi.org/10.1016/j.schres.2017.10.010 PMID:29055568

23. Saito T, Kondo K, Iwayama Y, Shimasaki A, Aleksic B, Yamada K, Toyota T, Hattori E, Esaki K, Ujike H, Inada T, Kunugi H, Kato T, et al. Replication and cross-phenotype study based upon schizophrenia GWASs data in the Japanese population: support for association of MHC region with psychosis. Am J Med Genet B Neuropsychiatr Genet. 2014; 165B:421–27. https://doi.org/10.1002/ajmg.b.32246 PMID:24888570

24. Ormel PR, van Mierlo HC, Litjens M, Strien MEV, Hol EM, Kahn RS, de Witte LD. Characterization of macrophages from schizophrenia patients. NPJ Schizophr. 2017; 3:41. https://doi.org/10.1038/s41537-017-0042-4 PMID:29138398

25. Andreassen OA, Harbo HF, Wang Y, Thompson WK, Schork AJ, Matingsdal M, Zuber V, Bettella F, Ripke S, Kelsoe JR, Kendler KS, O’Donovan MC, Sklar P, et al, and Psychiatric Genomics Consortium (PGC) Bipolar Disorder and Schizophrenia Work Groups, and International Multiple Sclerosis Genetics Consortium (IMSGC). Genetic pleiotropy between multiple sclerosis and schizophrenia but not bipolar disorder: differential involvement of immune-related gene loci. Mol Psychiatry. 2015; 20:207–14. https://doi.org/10.1038/mp.2013.195 PMID:24468824

26. Hargreaves A, Anney R, O’Dushlaine C, Nicodemos KK, Gill M, Corvin A, Morris D, Donohoe G, and Schizophrenia Psychiatric Genome-Wide Association Study Consortium (PGC-SCZ), and Wellcome Trust Case Control Consortium 2. The one and the many: effects of the cell adhesion molecule pathway on neuropsychological function in psychosis. Psychol Med. 2014; 44:2177–87. https://doi.org/10.1017/S0033291713002663 PMID:24284030

27. Mohammadi A, Rashidi E, Amooeian VG. Brain, blood, cerebrospinal fluid, and serum biomarkers in schizophrenia. Psychiatry Res. 2018; 265:25–38. https://doi.org/10.1016/j.psychres.2018.04.036 PMID:29680514

28. Morgan LZ, Rollins B, Sequiera A, Byerley W, DeLisi LE, Schatzberg AF, Barchas JD, Myers RM, Watson SJ, Akil
H, Bunney WE Jr, Vawter MP. Quantitative Trait Locus and Brain Expression of HLA-DPA1 Offers Evidence of Shared Immune Alterations in Psychiatric Disorders. Microarrays (Basel). 2016; 5:6. https://doi.org/10.3390/microarrays5010006 PMID:26998349

29. Riedel M, Spellmann I, Schwarz MJ, Strassnig M, Sikorski C, Möller HJ, Müller N. Decreased T cellular immune response in schizophrenic patients. J Psychiatr Res. 2007; 41:3–7. https://doi.org/10.1016/j.jpsychires.2005.11.007 PMID:16434055

30. Tanaka KF, Shintani F, Fujii Y, Yagi G, Asai M. Serum interleukin-18 levels are elevated in schizophrenia. Psychiatry Res. 2000; 96:75–80. https://doi.org/10.1016/S0165-1781(00)00196-7 PMID:10980328

31. Mudge J, Miller NA, Khrebtukova I, Lindquist IE, May GD, Huntley JJ, Luo S, Zhang L, van Velkinburgh JC, Farmer AD, Lewis S, Beavis WD, Schilkey FD, et al. Genomic convergence analysis of schizophrenia: mRNA sequencing reveals altered synaptic vesicular transport in post-mortem cerebellum. PLoS One. 2008; 3:e3625. https://doi.org/10.1371/journal.pone.0003625 PMID:18985160

32. Devor A, Andreassen OA, Wang Y, Mäki-Marttunen T, Smeland OB, Fan CC, Schork AJ, Holland D, Thompson WK, Witoelar A, Chen CH, Desikan RS, McEvoy LK, et al. Genetic evidence for role of integration of fast and slow neurotransmission in schizophrenia. Mol Psychiatry. 2017; 22:792–801. https://doi.org/10.1038/mp.2017.33 PMID:28348379

33. Gnanapavan S, Giovannoni G. Neural cell adhesion molecules in brain plasticity and disease. Mult Scler Relat Disord. 2013; 2:13–20. https://doi.org/10.1016/j.msard.2012.08.002 PMID:25877450

34. van Kammen DP, Poltorak M, Kelley ME, Yao JK, Gurklis JA, Peters JL, Hemperly JJ, Wright RD, Freed WJ. Further studies of elevated cerebrospinal fluid neuronal cell adhesion molecule in schizophrenia. Biol Psychiatry. 1998; 43:680–86. https://doi.org/10.1016/S0006-3223(97)00324-7 PMID:9583002

35. Nguyen TT, Dev SI, Chen G, Liou SC, Martin AS, Irwin MR, Carroll JE, Tu X, Jeste DV, Eyler LT. Abnormal levels of vascular endothelial biomarkers in schizophrenia. Eur Arch Psychiatry Clin Neurosci. 2018; 268:849–60. https://doi.org/10.1007/s00406-017-0842-6 PMID:28942562

36. Honer WG, Falkai P, Young C, Wang T, Xie J, Bonner J, Hu L, Boullianne GL, Luo Z, Trimble WS. Cingulate cortex synaptic terminal proteins and neural cell adhesion molecule in schizophrenia. Neuroscience. 1997; 78:99–110. https://doi.org/10.1016/S0306-4522(96)00489-7 PMID:9135092

37. Aberg KA, Liu Y, Bukszár J, McClay JL, Khachane AN, Andreassen OA, Blackwood D, Corvin A, Djurovic S, Gurling H, Ophoff R, Pato CN, Pato MT, et al. A comprehensive family-based replication study of schizophrenia genes. JAMA Psychiatry. 2013; 70:573–81. https://doi.org/10.1001/jamapsychiatry.2013.288 PMID:23894747

38. Smith MJ, Wang L, Cronenwett W, Goldman MB, Mamah D, Barch DM, Csernansky JG. Alcohol use disorders contribute to hippocampal and subcortical shape differences in schizophrenia. Schizophr Res. 2011; 131:174–83. https://doi.org/10.1016/j.schres.2011.05.014 PMID:21658914

39. Bergink V, Gibney SM, Drexhage HA. Autoimmunity, inflammation, and psychosis: a search for peripheral markers. Biol Psychiatry. 2014; 75:324–31. https://doi.org/10.1016/j.biopsych.2013.09.037 PMID:24286760

40. Strous RD, Shoenfeld Y. Schizophrenia, autoimmunity and immune system dysregulation: a comprehensive model updated and revisited. J Autoimmun. 2006; 27:71–80. https://doi.org/10.1016/j.jaut.2006.07.006 PMID:16997531

41. Ermakov EA, Ivanova SA, Buneva VN, Nevinsky GA. Hydrolysis by catalytic IgGs of microRNA specific for patients with schizophrenia. IUBMB Life. 2018; 70:153–164. https://doi.org/10.1002/iub.1712 PMID:29343194

42. Ruderfer DM, Ripke S, McQuillin A, Boocock J, Stahl EA, Pavlides JM, Mullins N, Charney AW, Ori AP, Looi LM, Domenici E, Di Fiorio A, Papiol S, et al, and Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium. Electronic address: douglas.ruderfer@vanderbilt.edu, and Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium. Genomic Dissection of Bipolar Disorder and...
43. Mathelier A, Shi W, Wasserman WW. Identification of altered cis-regulatory elements in human disease. Trends Genet. 2015; 31:67–76. https://doi.org/10.1016/j.tig.2014.12.003 PMID:25637093

44. Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. G3 (Bethesda). 2011; 1:457–70. https://doi.org/10.1534/g3.111.001198 PMID:22384356

45. Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. Nat Methods. 2013; 10:5–6. https://doi.org/10.1038/nmeth.2307 PMID:23269371

46. Guo L, Wang J. rsSNPBase 3.0: an updated database of SNP-related regulatory elements, element-gene pairs and SNP-based gene regulatory networks. Nucleic Acids Res. 2018; 46:D1111–16. https://doi.org/10.1093/nar/gkx1101 PMID:29140525

47. Li T, Underhill J, Liu XH, Sham PC, Donaldson P, Murray RM, Wright P, Collier DA. Transmission disequilibrium analysis of HLA class II DRB1, DQA1, DQB1 and DPB1 polymorphisms in schizophrenia using family trios from a Han Chinese population. Schizophr Res. 2001; 49:73–78. https://doi.org/10.1016/S0920-9964(00)00111-0 PMID:11343866